Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Air Force Personnel and Training Research Center

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *AF Personnel and Training Research Center*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
NEAL E. MILLER, *Yale University*
KENNETH W. SPENCE, *State University of Iowa*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Assistant Managing Editor
Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS

CONTENTS

Goal-Box Cues and Pattern of Reinforcement: D. FEHRER 361

An Analysis of Tracking Behavior in Terms of Lead-Lag Errors: W. D. GARVEY AND L. L. MITNICK 372

Learning and Performance in a Complex Tracking Task as a Function of Visual Noise: G. E. BRIGGS, P. M. FITTS, AND H. P. BAHRICK 379

Perception of Nonsense Passages in Relation to Amount of Information and Speech-to-Noise Ratio: I. MILLER 388

Rest and Warm-Up in Bilateral Transfer on a Pursuit Rotor Task: L. C. WALKER, C. B. DESOTO, AND M. W. SHELLY 394

Conflict and Information-Theory Variables as Determinants of Human Perceptual Curiosity: D. E. BERLYNE 399

The Vertical-Horizontal Illusion and the Visual Field: T. M. KÜNNAPAS 405

Differential Eyelid Conditioning as a Function of Stimulus Similarity and Strength of Response to the CS: M. D. GUNTHER 408

Contingent Partial Reinforcement and the Anticipation of Correct Alternatives: H. BRAND, J. M. SAKODA, AND P. J. WOODS 417

American Psychological Association

Vol. 53 No. 6 June 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, P. O. Box 1557, Lackland Air Force Base, Texas. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 53, NO. 6 JUNE, 1957

GOAL-BOX CUES AND PATTERN OF REINFORCEMENT[1]

DAVID FREIDES[2]

Yale University

When the alley-running response of the rat is to be studied under conditions of partial reinforcement, it is necessary to structure the situation so that *S* may not receive discriminatory cues before the running response is completed. The *S* cannot be allowed to perceive that the reinforcement is not present until it has arrived in the goal box. Researchers therefore arrange curtains and/or turns at the end of the runway to conceal visual cues and they provide the same olfactory cues on all trials. Of necessity the cue conditions of the goal box vary with the presence or absence of reinforcement. However, in many definitive studies (**8, 9, 10**), certain goal-box conditions are deliberately varied with the reinforcement schedule in order to minimize the effects of secondary reinforcement. The effect of such cue variations has not been examined nor has goal-box behavior been studied directly. It was the purpose of the present study systematically to vary the goal-box cues and the pattern of reinforcement and to measure response strength in both the alley and the goal box in an attempt to clarify some of the characteristics of the partially reinforced response, particularly its resistance to extinction.

METHOD

Design.—The design of the study (Table 1) involved a comparison of response strength in the alley and in the goal box when (*a*) goal-box conditions were maximally similar on both reinforced and nonreinforced trials, and (*b*) when they were strikingly differentiated. Provision was made for comparisons with responses under continuous reinforcement.

Similarity of goal-box conditions was manipulated by employing black and white goal boxes. If the conditions of training were to be made maximally similar, the same colored goal box was employed on all training trials (1 cue). If the conditions of training were to be differentiated, goal boxes of both colors were employed (2 cue).

Two levels of consistency of reinforcement, 50% and 100%, and two levels of similarity of goal-box conditions, 1 cue and 2 cue, were employed during acquisition in a factorial design. Half the *Ss* trained under each of the four resulting combinations were extinguished under one of the goal-box conditions encountered in training. Where only one goal-box condition was encountered in training (1 cue), the other half of the *Ss* were extinguished under novel goal-box conditions. Color per se was balanced

[1] This report is based on a dissertation presented for the degree of Doctor of Philosophy in Yale University, June 1956. The author is deeply indebted to Dr. F. A. Logan, chairman of the thesis committee, and Drs. N. E. Miller and K. C. Montgomery, members, for their advice and encouragement.

[2] Now at Army Medical Service School, Ft. Sam Houston, Texas.

TABLE 1

GOAL BOX CONDITIONS DURING ACQUISITION, EXTINCTION AND CUE REVERSAL EXTINCTION

(B and W refer to black and white goal boxes. The subscripts refer to the history of the *S*'s experience in that goal box: Reinforcement, (+); nonreinforcement, (−); no experience, (0). Only one of the color combinations is presented. The actual colors were reversed for one-half the *S*s in each group. *N* refers to the total *N* in each group. For explanation of other terms and the uneven *N*, see text.)

Condition	Acquisition Reinforcement Schedule						
	100%			50%			
	Acquisition Similarity			Acquisition Similarity			
	1 Cue		2 Cue	1 Cue		2 Cue	
	Gp. I (N = 14)	Gp. II (N = 13)	Gp. III (N = 13)	Gp. IV (N = 13)	Gp. V (N = 14)	Gp. VI (N = 13)	Gp. VII (N = 14)
Acquisition reinforcement	B_+	B_+	B_+ W_+	B_+	B_+	B_+	B_+
Acquisition nonreinforcement				B_-	B_-	W_-	W_-
Extinction	B_+	W_0	B_+	$B_\pm$	W_0	B_+	W_-
Cue reversal extinction	W_0	B_+	W_+	W_0	$B_\pm$	W_-	B_+

within groups, i.e., within each group black or white was designated as the cue associated with reinforcement. Thus two extinction groups resulted from all but one of the training conditions. Only one extinction group was designated in the case of *S*s trained with 100% reinforcement under differentiated conditions inasmuch as the subdivisions for extinction were functionally equivalent, differing only in the color of goal-box cues. Additional data were obtained by continuing extinction under reversed stimulus conditions. For instance, a group originally extinguished to the condition associated with reinforcement (Group VI) was switched to the condition associated with nonreinforcement. This phase of the experiment is termed Cue-Reversal Extinction (CRE).

Subjects.—The *S*s were 94 experimentally naive, female, black hooded rats, 90 to 120 days old at the beginning of the experiment. They were obtained from three different commercial sources but were homogeneous as to source within replications.

Apparatus.—The apparatus shown in Fig. 1 was constructed of ¼-in. plywood. The alley, painted an intermediate gray throughout, was 10 in. high and was covered with wire mesh. Metal guillotine type doors were of the same color. The floor panels, hinged at the side nearest the start box, were flush with the rest of the floor and activated microswitches when depressed. Illumination was provided by light bulbs (7½ w.) which were mounted on a frame 16 in. above the floor, over the start box, runway (at points ⅓ and ⅔ the length of the runway), right angle turn, and goal box. Two thicknesses of cheesecloth were draped over the frame so as to minimize extraneous visual cues.

The goal boxes contained water cups ½ in. in diameter and 1¼ in. above the floor. A solid, ½ in. thick wood barrier, extending the width of the goal box and rising ½ in. above the level of the water cup was mounted against the front of the water cup. This barrier served to conceal the presence or absence of reinforcement until *S* "nosed" the water cup or at least moved very close to it. All wood parts and the water cups were painted flat black in two goal boxes and flat white in the other two. The boxes were further differentiated by brass sheeting stapled to the floor of the white goal boxes and wire mesh stapled to the floor of the black goal boxes.

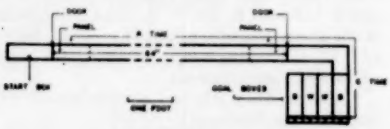

FIG. 1. Apparatus.

Metallic floors and a strip of brass sheeting across the top of the barrier constituted an open switch for a low amp contact relay. Additional goal boxes, three without and two with barriers, were constructed for purposes of preliminary training and were painted gray throughout. Living cages measured 15 × 30 × 10 in. and housed six to nine *Ss*. Individual waiting cages, 4 × 10 × 5 in., were constructed of unpainted wood with mesh floors and covers.

Reinforcement, chosen in part to minimize or eliminate olfactory cues, consisted of two drops of 16% sucrose solution (prepared by dissolving 16 gm. commercial sugar in 84 gm. tap water). It was placed in the water cup with a common eye dropper through a hole drilled in the wall of the goal box 1 in. above the water cup. On any single day, reinforcement was administered in only one of the pair of goal boxes of the same color; the other goal box being used for nonreinforced trials. A single goal box was used for purposes of reinforcement generally on alternate days.

Three Standard Electric timers and all relays were placed in a Celotex box which was mounted on rubber within another Celotex box, thereby reducing sounds from the relays and clocks to a noise level below that present in the room. Raising the first guillotine door started the first clock, depressing the first floor panel started the second clock, and depressing the second floor panel simultaneously stopped the first two clocks and started a third clock. The third clock was stopped when *S* touched the brass stripping on the barrier just in front of the water cup. (In addition a manually operated switch was provided to stop the third clock in certain contingencies to be described below.) Thus the clocks were activated in such a way as to measure (*a*) start time plus time in the runway (T time), (*b*) time in the runway (R time), and (*c*) time of approach to the water cup in the goal box (G time).

The clock readings were transformed into reciprocals and multiplied by the distance run to provide the speed measure, in terms of which all results are presented. The actual points on the panels at which the microswitches were depressed was estimated by sample observations to be 5 in. past the beginning of the panel. Accordingly runway speeds are based on a distance of 6.17 ft., and goal box speeds on 2.73 ft.

Procedure.—Upon arrival at the laboratory, rats were randomly assigned to living cages and placed on an ad lib. diet of food and water for at least one week. During all phases of the experiment solid food was administered by strewing Purina checkers about the floors of the living cages. Preliminary training was preceded by deprivation of all solid food for two days. During each of the next seven days, *Ss* were placed in the waiting cages, weighed, returned to the waiting cages and then transferred back to their living cages where solid food was available for 1 hr. The *Ss* were familiarized with the water cups during the next five days. Water bottles were removed and *Ss* were allowed free access to dry food for 1 hr. after which they were removed to the waiting cages. The water cups of the preliminary training goal boxes, with barriers (X) and without barriers (Y), were filled with tap water. The *Ss* were placed first in Goal Box Y and then in Goal Box X and allowed to find and drink the water. For three days they were given 10 min. in which to drink. Successful *Ss* were placed only in Goal Box X on the last two days, while unsuccessful *Ss* continued to receive additional experience in Goal Box Y. Two consecutive responses in Goal Box X was a criterion by which *Ss* were selected for the experiment; seven *Ss* failed to meet this criterion. After completion of the preliminary training, *Ss* were assigned at random to the various experimental groups.

During the experiment proper, drive was maintained by allowing access to solid food for 45 min. after each daily session. In addition, water bottles were removed 4 hr. before each session. On each trial, *S* was placed in the start box and after 20 sec. *E* raised the start box door. After *S* passed the second guillotine door, it was lowered to prevent retracing. On rewarded trials *S* was removed to the waiting cage just after the sugar solution was consumed. (It generally took *Ss* about 10 sec. to consume the reinforcement.) On nonreinforced training trials and on all extinction trials, *S* was removed after 15 sec.

All *Ss* were given 120 spaced training trials, 60 extinction trials, and 30 cue-reversal extinction trials. A single trial was administered on each of the first 2 days; 3 trials were administered on the third day and 5 trials on the fourth day. Thereafter, 10 trials a day were administered with a minimum intertrial interval of 15 min. The running order within replications was randomly varied from day to day. For *Ss* receiving partial reinforcement, the order of reinforced and nonreinforced trials was the same on any particular day but varied from day to day. On the first two days a single reinforced trial was administered. The order for the third day was + − +, for the fourth day + − + − +, and for the fifth day + + − + − − + − − +. For the *last* 10 days of training, 10 different orders were chosen from among those published by Gellerman (**2, 3**). Of the first 10 trials, 70% were reinforced; thereafter 50% of the trials were reinforced.[3]

[3] If an animal was lost during one period, it was replaced by adding another to one of the

During extinction a trial was terminated if *S* did not leave the start box in 60 sec., and only a T time of 60 was recorded. If *S* left the start box within 60 sec. but did not reach the second floor panel within another 60 sec., the trial was terminated and T time and R time were both recorded as 60. During all phases of the experiment, if, after depressing the second panel, *S* failed to reach the food cup in less than 15 sec., a G time of 15 was recorded. (In those few instances when *S* reached the second floor panel but then reversed and depressed the first panel a second time before proceeding to the goal box, the first T and R times were recorded, the G clock was reset, and the last G time was recorded.)

Early in the study some *S*s were observed to move their heads back and forth over the water cup without touching the barrier and stopping the G time clock. Since this clock was intended to measure approach time to the water cup, and since the behavior described occurred at the water cup without stopping the clock, the G time clock was stopped manually whenever it was apparent that *S* was in a position to perceive the contents of the water cup. All *S*s were closely observed, but only those *S*s receiving partial reinforcement under 1-cue conditions (Groups IV and V) displayed this behavior (on approximately 25% of negative trials at the end of training), and these data were specifically identified. Every effort was made to use the manual switch conservatively, hence these data tend to overestimate G time.

replications of the succeeding period. Four animals died of unknown causes after training began and five were discarded for failure to run. The criterion employed was four or more trials in which *S* failed to leave the start box in 60 sec. on Days 6 and 7 of training. Four *S*s were lost during the last period, each from a different group; hence the unequal *N*.

Results

The results were tested statistically by analysis of variance whenever the assumption of homogeneity of variance could not be rejected by means of Bartlett's test. Acquisition asymptotes were based on performance during the last five days of training.

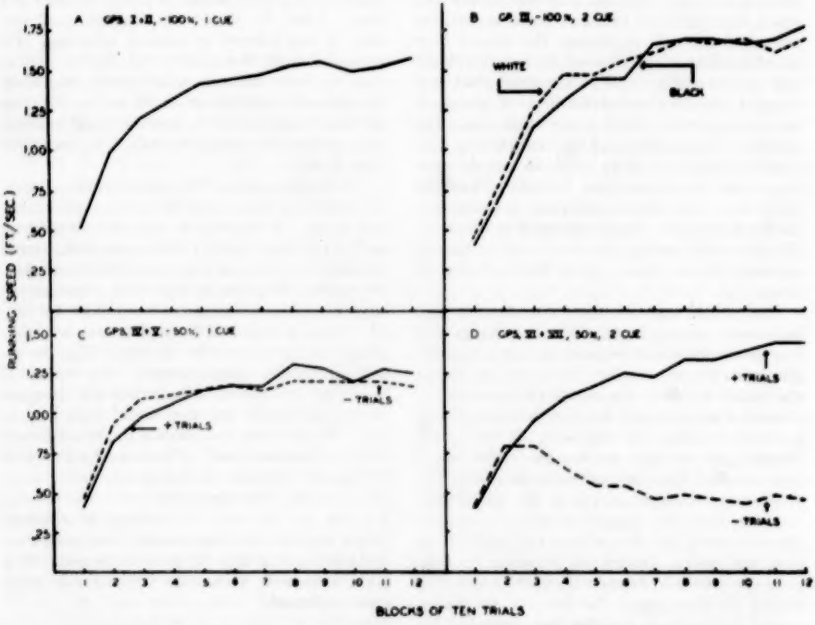

FIG. 2. Mean goal-box speed under four conditions of acquisition.

TABLE 2

COMPARISONS OF ASYMPTOTIC GOAL-BOX SPEEDS ON REINFORCED TRIALS

Group	*N*	Mean Speed	*t* of Difference		
			A	B	C
A. 100%, 1 cue	27	1.54			
B. 100%, 2 cue	13	1.68	1.16		
C. 50%, 1 cue	27	1.26	2.80*	3.42*	
D. 50%, 2 cue	27	1.39	1.50	2.37**	1.31

* $P < .01$.
** $.02 < P < .05$.

Goal-box performance.—Performance in the goal box under the four conditions of acquisition is shown in Fig. 2. At asymptote, *Ss* given continuous reinforcement under differentiated conditions (Fig. 2B) responded about as rapidly to one cue as to the other (asymptotic mean diff. = .0371, $t = 1.209$, $P = .25$); *Ss* given partial reinforcement under differentiated conditions (Fig. 2D) responded discriminatively on positive and negative trials (asymptotic mean diff. = .9230, $t = 18.148$, $P < .0001$); and *Ss* given partial reinforcement under constant conditions responded about as rapidly on negative as on positive trials (asymptotic mean diff. = .0666, $t = 2.259$, $.02 < P < .05$).[4]

Comparisons among asymptotic goal-box speeds on *reinforced* trials were tested by means of analysis of variance and yielded an F of 4.90 (3 and 90 *df*, $P < .01$).[5] Table 2 presents the results of the subsequent t tests of the differences between the means. It is apparent that at asymptote, *Ss* given 100% reinforcement approached the water cup more rapidly than *Ss* given 50% reinforcement under constant conditions. The *Ss* given 50% reinforcement under differentiated conditions approached the water cup at an intermediate speed.

Performance during extinction for the five groups running under *familiar* conditions is shown in Fig. 3. It is apparent that the group running to the negative cue (VII) maintained throughout extinction the same low speed of approach to the water cup that was evident at the end of training (difference between mean asymptotic speed and mean extinction speed = .003), while the four groups running to cues which elicited rapid approach speeds at the end of acquisition reversed their relative positions during extinction, i.e., the fastest groups appeared to be least resistant to extinction.

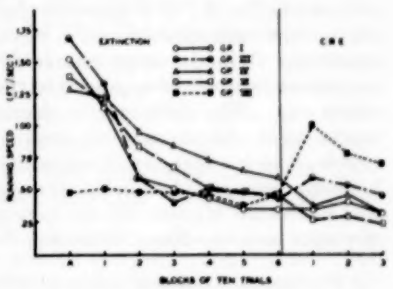

FIG. 3. Mean goal-box speed during extinction and cue-reversal extinction (CRE) for groups running under familiar conditions. The initial points (A) are mean asymptotic acquisition speeds. Groups running under novel conditions (II and V) are shown in Fig. 4.

[4] The significance of this small difference was attributable to the ability of these *Ss* to discriminate the presence or absence of reinforcement while at the water cup without touching the barrier, and the consequent delay entailed in stopping the clocks manually on negative trials. (See Procedure.) The more reliable estimate of these *Ss*' asymptotic acquisition level, speed on reinforced trials, was employed for all subsequent comparisons.

[5] Color per se, and the group color interaction, yielded insignificant F's in this and all other analyses of extinction performance. Consequently any variance attributable to color was included in the error estimate and the *df*'s were reapportioned accordingly.

For statistical analysis of the extinction results of the four groups showing rapid approach, a "drop" score was computed for each *S* separately by subtracting the mean of all 60 extinction trials from the mean asymptotic acquisition score.[6] The greatest drops occurred in the 1- and 2-cue 100% groups (I and III), the least in the 1-cue partially reinforced group (IV). The group running to the positive cue (VI) was intermediate. Since the variances of these "drop" scores could not be assumed to be homogeneous (Bartlett's test, $\chi^2 = 14.42$, 3 *df*, $P < .01$), the differences were tested, as appropriate because of *N*, by means of nonparametric tests, Williamson's *T* (6) and the Mann-Whitney *U* test (6). All differences except that between the two 100% groups were significant beyond the .01 level.

Performance during extinction of the two groups running to novel conditions and of the groups trained under the same conditions and extinguished under conditions of maximal similarity to those of training is shown in Fig. 4. It is apparent that the groups extinguished under novel conditions showed a large immediate decrement in speed of approach to the water cup. The difference in decrements from the asymptotic acquisition level to the first day of extinction between groups identically trained was evaluated statistically by means of *t* tests and was found to be significant beyond the .01 level in each case.

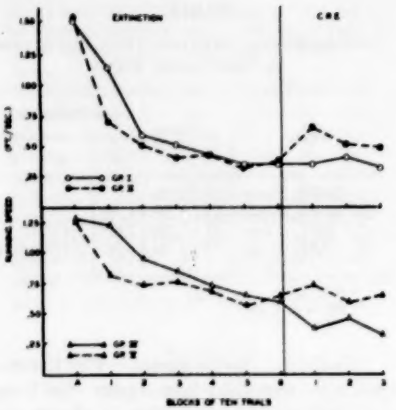

FIG. 4. Mean goal-box speed during extinction and cue-reversal extinction (CRE) for groups running to novel conditions (Group II and V). Groups trained under the same conditions and extinguished under conditions maximally similar to those employed during acquisition are shown for comparative purposes. The initial points (A) are mean asymptotic acquisition speeds.

To compare extinction of approach to the water cup under novel conditions subsequent to the initial drop, a score was computed by subtracting mean performance on the last day of extinction from mean performance on the first day of extinction. The groups running to novel cues showed significant ($P < .02$) drops in speed of approach to the water cup. However, these drops were significantly smaller ($P < .01$) than the drops shown by all other groups with the exception of the group running to the negative cue. This group (VII) did not show a significant drop between the first and last days of extinction. From these relationships it is apparent that most of the extinction of approach to the water cup for *S*s responding in the presence of novel cues occurred with the change from fa-

[6] This measure is presented since a straightforward covariance analysis did not appear entirely justified. (The variances of the mean extinction scores could not be assumed to be homogeneous.) However, the analysis of covariance yielded an *F* of 3.79 (3 and 49 *df*, $.01 < P < .05$) and the ranking of the mean regressed goal-box extinction speeds of individual groups, from slowest to fastest, was as follows (dashes indicate significant differences between adjacent groups): – III, I, – VI, – IV,

miliar to novel cues and that relatively little extinction occurred subsequently.

Performance during cue reversal extinction is shown in Fig. 3 and 4. Mean differences of goal-box speeds on the first day of cue-reversal extinction and the last day of extinction, and the significance of the differences, are presented in Table 3. It is apparent that *S*s encountering novel cues (I and IV) and *S*s encountering negative cues (VI) in the goal box displayed decrements in speed of approach to the water cup, while *S*s encountering cues previously associated with reinforcement displayed increments in speed of approach to the water cup. By far the greatest increment occurs in Group VII where the positive cue is re-encountered.

Runway performance.—In contrast to the results obtained in the goal box, asymptotic runway speeds reached under the four conditions of acquisition did not differ significantly. Although the trend is for the partial reinforcement groups to run slower than the continuous reinforcement groups, analysis of variance yielded an *F* of only .85.

TABLE 3

MEAN DIFFERENCE AND PROBABILITY, FIRST DAY OF CUE REVERSAL EXTINCTION MINUS LAST DAY OF EXTINCTION, FOR GOAL-BOX AND RUNWAY SPEED

Gp.	Goal-Box Speed		Runway Speed	
	Mean Diff.	*P*	Mean Diff.	*P*
I	− .006	.42	.265	.05
II	.227	.01	.339	.01
III	.236	.058	.105	.68
IV	− .184	.012	.401	.04
V	.078	.08	.296	.02
VI	− .191	.01	.080	.31
VII	.557	.01	.574	.01

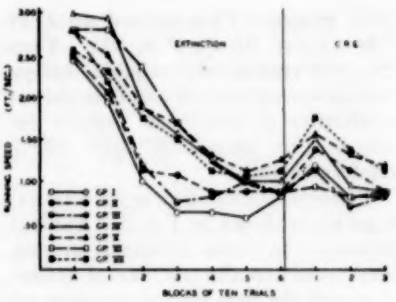

FIG. 5. Mean runway speed during extinction and cue-reversal extinction (CRE) for each of the seven conditions of extinction. The initial points (A) are mean asymptotic acquisition speeds.

Performance during extinction is shown in Fig. 5. A "drop" score was calculated for each *S* separately by subtracting the mean of 60 extinction trials from the asymptotic acquisition score.[7] An analysis of variance of these "drop" scores yielded an insignificant *F* of 1.74. However, analysis of variance of the mean drop over the first 30 extinction trials yielded an *F* of 5.49 (6 and 87 *df*, $P < .01$). Subsequent *t* tests revealed that the only significant individual comparisons were between partially and continuously reinforced groups. The drop scores of all 60 extinction trials were then reanalyzed, combining the data into two groups: the 100% groups and the

[7] This measure is presented since a covariance analysis did not appear entirely justified. (The variances of the mean runway extinction scores could not be assumed to be homogeneous.) However, the analysis of covariance yielded an *F* of 3.04 (6 and 86 *df*, $P < .01$) and the ranking of the mean regressed runway extinction speeds of individual groups, from slowest to fastest, was as follows (dashes indicate significant differences between adjacent groups): − I, III, II, − V, VII, VI. Group IV was intermediate between Groups II and V and differed significantly from neither.

50% groups. This yielded an F of 7.24 (1 and 92 df, $P < .01$). Thus the one statistically reliable finding for runway extinction was that partial reinforcement produced greater resistance to extinction than 100% reinforcement.

Performance during cue-reversal extinction is shown in Fig. 5. Table 3 presents the mean difference in runway speeds on the last day of extinction and the first day of cue-reversal extinction. All groups but the one running to the negative cue (VI) and the one trained with 100% reinforcement under differentiated conditions (III) show statistically significant increases in runway speed. The CRE results on runway performance contrast with those on goal-box performance in that (*a*) *S*s running to novel stimulation (I and IV) show decrements in goal-box speed but show significant increments in runway speed; (*b*) *S*s trained with 100% reinforcement under differentiated conditions (III) which are extinguished first to one condition and then, in cue-reversal extinction, to the other, show at the time of reversal a significant rise in goal-box speed but only an insignificant rise in runway speed. In other cases, significant increments in goal-box speed are accompanied by significant increments in runway speed.

These analyses took into account only trials actually run, i.e., trials terminated when *S* had not entered the runway in 60 sec. were not considered. Re-analysis of runway speed data incorporating minimal values (1/60 sec.) for every trial in which the runway was not entered produced results comparable to those presented above. Start-plus-runway speed yielded comparable but more variable and therefore less reliable results than runway speed, and are also not presented (**2**).

Discussion

The 1 cue-2 cue dimension was systematically varied to provide controls for the procedure followed by Sheffield (**8**), Weinstock (**9**), and Wilson, Weiss, and Amsel (**10**) in their studies of continuous and partial reinforcement. These *E*s removed a perceptually conspicuous food cup from the goal box on all nonreinforced trials. An experiment by Elam, Tyler, and Bitterman (**1**) indicated that resistance to extinction diminished significantly following 50% 2-cue training if the goal box contained the positive rather than the negative stimulus during extinction. On the basis of the Elam study, it was hypothesized that the typical finding of greater resistance to extinction following partial as compared with continuous reinforcement (**4**) might have been due to the incidentally varied characteristics of the goal box rather than to the schedule of reinforcement per se. On the other hand, a study by Notterman (**7**) comparable to that of Elam, Tyler, and Bitterman (**1**) cast doubt on this possibility. The present experiment encompassed within a single design the conditions of all the above studies, plus additional control groups as dictated by a factorial design.

The finding of no significant difference in runway speed during extinction between Groups VI and VII is comparable to the result obtained by Notterman (**7**) and is not in accord with the result obtained by Elam, Tyler, and Bitterman (**1**). The runway data during extinction yielded statistically significant differences between partially reinforced and continuously reinforced groups. This evidence is in accord with that provided by Wilson, Weiss, and Amsel (**10**) and Weinstock (**9**) and is contrary to the results obtained by Sheffield (with identically spaced intervals) in an earlier and widely cited experiment (**8**).

When the goal-box data are taken into consideration, the significant empirical differences in the runway data appear somewhat perplexing. Considered alone, the goal-box data suggest that behavior in the goal box varies in an orderly way as a function of both the schedule of reinforcement and the specific types of stimulation present in the goal box (positive cues, negative cues, novel cues, etc.). Runway behavior is generally assumed to be a consequence of events which occur in the goal box and "work their way back" into the runway. With this principle in mind, some of the runway-goal box relationships appear to be complementary and offer no special problem. For instance, relatively rapid extinction in the goal box is associated with relatively rapid extinction in the runway in the cases of Groups I and III (100% training, extinction in the presence of familiar cues). Similarly, the relatively flat goal-box extinction curves of all partially reinforced groups are associated with runway extinction curves which are relatively flat when compared with the curves of continuously reinforced groups. It is, however, also evident that the partially reinforced groups differ significantly among themselves in goal-box behavior, whereas they are not significantly different in the runway. Furthermore, trends in the extinction data, and more conspicuously in the cue-reversal extinction data, indicate that when novel cues are present in the goal box, runway responses are relatively more resistant to extinction while goal-box responses extinguish more rapidly.

Considerations of space preclude a detailed theoretical interpretation of these data here (see **2**). However, the goal-box data strongly suggest that a simple formulation concerning the effects of secondary reinforcement (e.g., cues associated with reinforcement, when present during extinction, serve to prolong extinction more than novel stimuli or cues associated with nonreinforcement) is untenable. What appears to be required is a detailed analysis of the responses elicited by each "type" of stimulus and a careful delineation of both the excitatory qualities of that stimulus and its inhibitory susceptibilities. For example, an analysis of extinction behavior in Group VI (50% 2-cue training, positive stimulus during extinction) needs to take into consideration not only the excitatory power of the positive stimulus but also such features as the following: (*a*) the stimulus has been continuously associated with reinforcement; (*b*) the organism has experienced nonreinforcement and has learned discriminatory responses to positive and negative stimuli (see acquisition data); (*c*) during extinction, the strong responses evoked by the excitatory powers of the stimulus are more likely to extinguish than weaker responses would be. In the case of Group VII (50% 2-cue training, negative stimulus during extinction), *a* above becomes—the stimulus has been continuously associated with nonreinforcement, and *c* becomes—under conditions of extinction, the weak responses evoked by the inhibitory stimulus are less likely to extinguish than strong responses would be.

Such an approach would ultimately take into consideration a wide array of variables. The goal-box data bring into focus but a few such variables. A number of other factors requiring investigation are suggested by the lack of complementary relationships between goal-box and runway data in this particular situation. Noteworthy among these are the length of the runway, i.e., the distance over which goal-box events must "work back," and the mechanism of cue concealment, whether right-angle turn (**7, 9**) jump gap (**1**), or one-way doors.

The goal-box data do not appear to support the theoretical interpretations of Weinstock (**9**), Wilson, Weiss, and Amsel (**10**), and Lauer and Estes (**5**). These writers maintain that increased resistance to extinction following partial reinforcement is a consequence of the

"habituation" (9) or "conditioning" (5, 10) of the running response to stimulation resulting from unconditioned responses to nonreinforcement or frustration. According to this position, continuously reinforced *Ss* do not acquire this type of S-R link and their running responses are readily disrupted during extinction by unconditioned responses to nonreinforcement. In other words, it is assumed that as partial-reinforcement training progresses, responses on positive and negative trials become more and more alike. The acquisition goal-box data of the present study provide evidence that behavior on reinforced and nonreinforced trials varies enormously. Where differential distance cues are provided, as in Groups VI and VII, a clear-cut discrimination developed. Where an attempt was made to keep differential cues to a minimum, rats gave strong indication of discriminatory behavior at the earliest possible moment (see last paragraph on procedure).

Summary

The purpose of this experiment was to observe the effects of variations in goal-box cues and pattern of reinforcement on the alley-running response of the rat during acquisition and extinction under conditions of distributed practice. Measurements of running speed in the goal box provided the means for studying response variations in the presence of different types of stimuli while measurements in the runway provided the means for studying the effects of such stimuli on antecedent responses. Stimuli were manipulated by using black and white goal boxes.

Ninety-four rats were trained to run down an alley and drink two drops of 16% sugar solution. Four training conditions (120 trials) involved two levels of consistency of reinforcement (50% and 100%) and two levels of goal-box similarity (1 cue and 2 cues). In the 50%, 2-cue condition, one cue was always presented on reinforced trials, the other on nonreinforced trials. The *Ss* trained with two cues were divided for extinction: half were run to each cue. Half the *Ss* trained under one cue were extinguished to the same cue and half to a novel cue. After 60 extinction trials the cues were reversed for 30 more extinction trials.

During acquisition, 50% and 100% reinforced *Ss* showed no significant asymptotic differences in runway speed. Asymptotic goal-box speeds varied as a function of both the pattern of reinforcement and goal-box cues, continuously reinforced *Ss* showing higher asymptotes than partially reinforced *Ss* and cues associated with reinforcement showing higher asymptotes than cues associated with nonreinforcement. Discriminatory responses were noted in the goal box where differential cues were first evident.

During extinction, partial reinforcement produced significantly greater resistance to extinction than continuous reinforcement in the runway. Goal-box cues were not significantly related to runway extinction speeds but were significantly related to goal-box extinction speeds.

The results are considered in the light of current conceptions of partial reinforcement, secondary reinforcement, and extinction.

References

1. Elam, G. B., Tyler, D. W., & Bitterman, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.
2. Freides, D. Goal box cues and pattern of reinforcement. Unpublished doctor's dissertation, Yale Univer., 1956.
3. Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 206–208.
4. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
5. Lauer, D. W., & Estes, W. K. Successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcement. *J. comp. physiol. Psychol.*, 1955, **48**, 8–13.
6. Moses, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.* 1952, **49**, 122–143.
7. Notterman, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary

reinforcement. *J. exp. Psychol.*, 1951, **41**, 161–169.

8. Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
9. Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.
10. Wilson, Wilma, Weiss, Elizabeth J., & Ambel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–60.

(Received May 23, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

AN ANALYSIS OF TRACKING BEHAVIOR IN TERMS OF LEAD–LAG ERRORS[1]

W. D. GARVEY AND L. L. MITNICK[2]

Naval Research Laboratory

One of the most useful approaches used in the study of human tracking behavior considers the human operator as an error-sensing and -reducing element in a continuous control system. The present study stems from a recent approach to the study of man-operated continuous control systems put forward by Birmingham and Taylor (1). Through the use of a mathematical model it is possible to compare the performance of the human operator in a compensatory tracking task with that of the mathematically simplest mechanism which might be substituted in the system to perform the operator's task. For example, it has been established that with a constant velocity input the replacement of the human tracker with the tracking mechanism shown in Fig. 1, which consists of two cascaded integrators with appropriate feed-forward loops, will reduce the tracking error to zero. If the input is a constant acceleration, an additional integrator with feed-forward loop must be added to the mechanism in Fig. 1, i.e., three cascaded integrators with appropriate feed-forward loops are required to perform the tracking task with zero error. In either case, if one integrator is removed from the mechanism, the output will lag the input; such lag will result in a constant tracking error. If two integrators are removed, a constantly increasing lag error will result. The purpose of the present study is to determine (*a*) the type of mechanism to which the operator's performance is analogous, i.e., if the operator's performance is analogous to a one-, a two-, or a three-integrator mechanism, and (*b*) if this analogy changes as a function of practice.

METHOD

Apparatus and procedure.—A simplified block diagram of the apparatus is shown in Fig. 2. The tracking task was compensatory, i.e., *Ss* sought to keep a dot of a cathode ray tube centered on a hairline by manipulating a handwheel. Tracking was only in one dimension, since the dot was free to move only in the

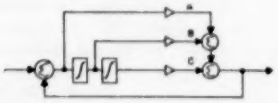

FIG. 1.[3] An example of the mathematically simplest mechanism which will reduce tracking error to zero when the input to a continuous control system is a constant rate. The system is shown to consist of two cascaded integrators with feed-forward loops around both. The circles containing sigmas represent differentials which add algebraically. The triangles represent amplifiers with appropriate gain settings. Path A represents the position component of the input; Path B, the rate component; and Path C, the acceleration component.

[1] The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department and the naval service at large.

[2] The authors are indebted to Dr. F. V. Taylor and Mr. H. P. Birmingham for their assistance in the formulation of the problem and their fruitful suggestions throughout the experiment. Also, they wish to thank Dr. W. B. Knowles for his assistance in setting up the apparatus.

[3] This figure and explanation are essentially taken from the work of Birmingham and Taylor (1), where examples of other similar mechanisms and more detailed discussion may be found.

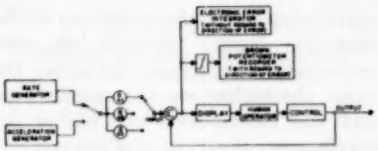

FIG. 2. Simplified block diagram of apparatus.

horizontal plane. The dot was forced off the hairline by either a velocity or an acceleration voltage input: the velocity inputs were 3.33 cm./sec., 6.67 cm./sec., and 13.33 cm./sec., and the acceleration inputs were .22 cm./sec.^{2}, .44 cm./sec.^{2}, and .88 cm./sec.^{2}. The handwheel exerted a positional control over the dot, i.e., a rotational change of the wheel produced a change in the position of the dot on the scope. The sensitivity of the handwheel, one revolution effecting a 7.2-cm. displacement of the target on the scope, was established after some preliminary experimentation so that the maximum rate which *Ss* had to turn the wheel to keep the dot centered did not exceed his handwheel-turning capacity.

Six naval enlisted men served as *Ss*. The general experimental design consisted of six *Ss*, six experimental conditions, and six 30-sec. trials counterbalanced in a $6 \times 6 \times 6$ square design. This design was permuted from day to day throughout the duration of the experiment, so that each *S* had six 30-sec. trials on each condition per day. The independent variables were (*a*) type of course input, three velocities and three accelerations, and (*b*) amount of practice. The dependent variables were (*a*) accuracy of performance measured in terms of integrated error scores, and (*b*) total lead- or lag-error scores accumulated during the trial. Both measures were taken only during the last 25 sec. of the 30-sec. trial. Due to the complexity of the record-reading task, the lead-lag error scores were recorded every third day. The integrated error scores were recorded each day, primarily to give *Ss* some knowledge of their performance, but only the data for every third day were used in the analysis. Before each trial *Ss* were told whether they were to track a constant rate or a constant acceleration of the target.

RESULTS

Average integrated error scores.—Performance accuracy, measured in terms of average integrated error, as a function of practice is shown in Fig. 3; the parameters are the three rate and three acceleration inputs. Using only the data for Days 19, 22, and 25, statistical tests (**7**) were made to determine the level of significance of the performance differences. The first column of Table 1 gives a summary of these tests. The results indicate that, in general, the faster either a rate or acceleration input, the poorer the performance. The differences between performance accuracy on slow rate and slow accele-

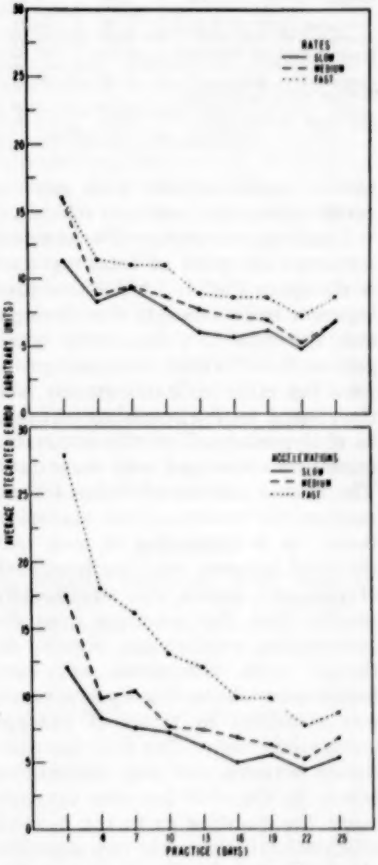

FIG. 3. Average integrated error as a function of practice. The parameters in the top graph are the three rate inputs; those in the lower graph are the three acceleration inputs.

TABLE 1
P VALUES OF PERFORMANCE DIFFERENCES MEASURED IN TERMS OF AVERAGE INTEGRATED ERROR AND TOTAL LAG-ERRORS

Comparison	*P* Values	
	Average Integrated Error	Total Lag-Errors
Slow rate vs. Medium rate	> .10	= .25
Slow rate vs. Fast rate	< .01	< .01
Medium rate vs. Fast rate	< .01	< .05
Slow accel. vs. Medium accel.	= .01	= .10
Slow accel. vs. Fast accel.	< .01	< .01
Medium accel. vs. Fast accel.	< .01	< .01
Slow accel. vs. Slow rate	> .05	< .01
Medium accel. vs. Medium rate	> .25	< .05
Fast accel. vs. Fast rate	> .05	< .05

ration, medium rate and medium acceleration, etc., were not reliable.

Total lag-error scores.—Performance measured in terms of total lag error is shown in Fig. 4. A positive total lag-error score indicates that throughout the trial *S*'s lag error totals greater than his lead error; a negative total lag error indicates greater lead than lag. A testing procedure identical to that performed on the integrated error scores was used with these data. The second column of Table 1 summarizes the results of the statistical tests. It is interesting to note that the total lag error resulting from each of the rate inputs was significantly smaller than that resulting from the comparable acceleration inputs, although such differences were not found to be reliable when performance was measured in terms of average integrated error. This finer discrimination between rate and acceleration inputs by the total lag-error measure raises the question as to the over-all relationship between the two measures of performance. Correlations were made between the two types of measures; the resulting *r*'s are shown in Table 2. The results show high correlation (with the exception of the slow-rate input) between the two measures. In general, the faster the input the higher are the correlations, and the correlations are higher for acceleration than the rate inputs.

The data used to plot the curves in Fig. 4 were tested (2) to determine if the lag or lead errors differed significantly from zero. The results of these tests are shown in Table 3, where a "+" indicates that the total lag-error score, averaged over all *S*s, differed significantly from zero, and

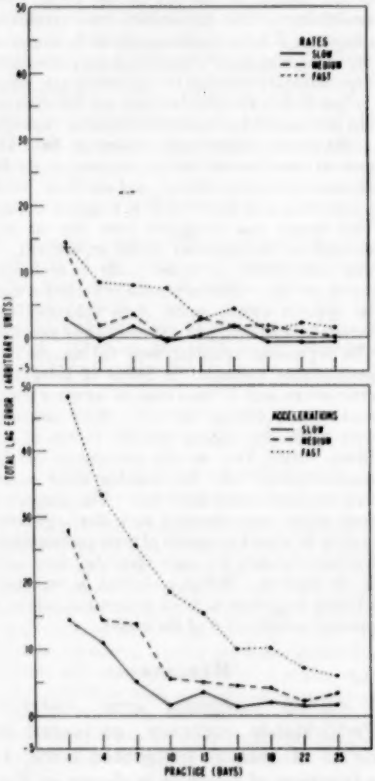

FIG. 4. Total lag error as a function of practice. The parameters in the top graph are the three rate inputs; those in the lower graph are the three acceleration inputs.

TABLE 2

OBTAINED CORRELATIONS (PEARSON *r*) BETWEEN AVERAGE INTEGRATED ERROR SCORES AND TOTAL LAG-ERROR SCORES

Category	*r*'s	
	Rate	Acceleration
Slow	+.031	+.693
Medium	+.532	+.895
Fast	+.608	+.888

"−" indicates that a significant lead error was obtained. The most apparent feature of Table 3 is the predominance of significant total lag error, i.e., only one of the 54 sets proves to have a significant total *lead* error. Another point of interest is the trend for fewer conditions to exhibit a significant total lag error as practice progresses. In the first stage of practice all but slow rate exhibits a significant total lag error, whereas in the last stage only the medium and fast accelerations show significant total lag error. It should be noted that the slow-rate condition shows a significant total *lead* error on the last trial.

Constancy of lag error during the trial.—To determine the constancy of the lag error during a tracking trial, each *S*'s trial was broken down into 5-sec. intervals. Then the difference between the score at each interval and its preceding interval was obtained. Since only the last 25 sec. of each trial was used, the 10-sec. point on the graph in Fig. 5 represents the average accumulation of lag error from the end of the first 5 sec. to the end of the second 5 sec.; the 15-sec. point represents the average accumulation of lag error between the end of the second 5 sec. and the end of the third 5 sec. Thus, if the curves are parallel to the abscissa, then the lag error accumulated from one period to another is constant. From such an analysis it may be established whether the rate of accumulation of lag error is constant, decreases, or increases during the tracking trial. To obtain some notion of the effect of practice on the forms of these curves, performance has been divided into three stages: (*a*) Day 1, (*b*) Day 13 (the middle of practice), and (*c*) Day 25 (the last day of practice). The results of this analysis are shown in Fig. 5. It should be remembered that Table 3 indicates if the total lead or lag error differed significantly from zero, whereas the asterisks in Fig. 5 indicate that the rate of build-up of the lag error changed significantly during the course of the trial. Thus, the over-all levels, not the trends, of the curves in Fig. 5 are the measures tested in Table 3. For example, the trend of the lag error for the fast acceleration on Day 25 was not found to be significant, but the total lag error, as seen in Table 3, was found to differ significantly from zero.

Statistical tests (**3**) of the trends of the curves in Fig. 5 indicate no significant increase or decrease in the

TABLE 3

STATISTICALLY SIGNIFICANT MEAN LEAD- OR LAG-ERROR SCORES FOR SIX CONDITIONS*

Days	Rate			Acceleration		
	Slow	Medium	Fast	Slow	Medium	Fast
1		+	+	+	+	+
4			+	+	+	+
7					+	+
10			+		+	+
13				+	+	+
16			+		+	+
19					+	+
22				+	+	+
25	−				+	+

* (+) indicates mean lag error significant at 5% level or better; (−) indicates mean lead error significant at 5% level or better.

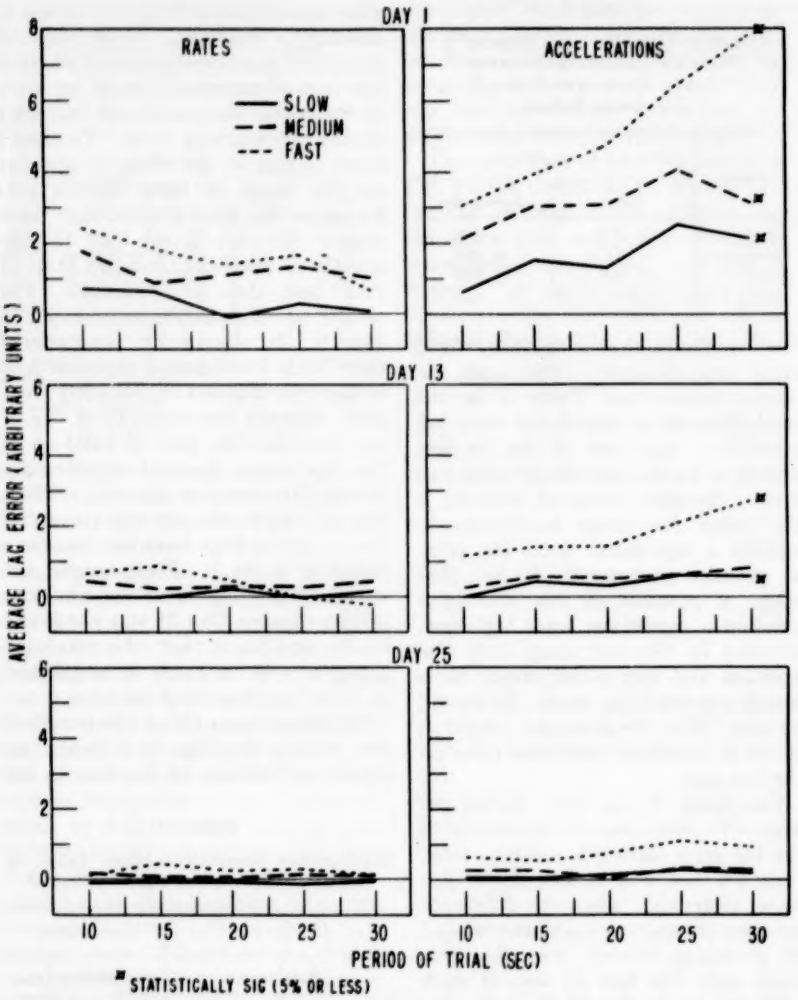

FIG. 5. Average lag error as a function of period of trial. The parameters in the three sets of graphs are the three rates and the three acceleration inputs. The top set of curves are the results from Day 1 of practice, the middle set are the results from Day 13, and the lower set are from Day 25.

accumulation of lag error during the course of the trial for any of the rate inputs during any of the stages of practice. In the case of the acceleration inputs the trends of all three curves were found to be significant on Day 1. At the middle stage of practice only the fast and slow acceleration curves show a significant trend. On the last day of practice *none* of the curves reveal a significant trend.

Discussion

The purpose of this study was to determine if tracking behavior of a human operator in a continuous control system can be analogized to the mathematically simplest mechanism which can be substituted to perform the operator's task; i.e., if an operator's behavior is analogous to a mathematical model consisting of one or more integrators. The results of the study indicate that the type of mechanism which may be substituted to provide performance analogous to that of the human operator differs as a function of the amount of practice the operator has had with the system. In general, at the beginning of practice the mechanism shown in Fig. 1 with one of the integrators removed may be substituted in the system used in this experiment to provide tracking performance equivalent to that of the human operator with either a rate or an acceleration input. This is attested by the fact that *S*s' performance showed constant lag errors on the medium- and fast-rate inputs and increasing lag errors on the slow-, medium-, and fast-acceleration inputs. At the end of the amount of practice imposed by this study *S* is performing analogously to the two-integrator mechanism shown in Fig. 1; for the data show that the lag errors are equivalent to zero for the medium- and fast-rate inputs and that there are constant lag errors for the medium- and fast-acceleration inputs. However, the results obtained from the slow-rate and slow-acceleration inputs are to some extent exceptions to these general statements. In the case of the slow-rate input, the human operator behaved analogously to a two-integrator mechanism from the first day of practice. Moreover, on the last day of practice there was a significant total *lead*-error score for the slow-rate input. There are several possible explanations for this result. It may have been due merely to chance, i.e., one significant total lead-error score in the 54 possible cases. It is also possible that there might have been an interaction effect among the conditions, i.e., *S* may have learned to apply a "correction factor" in order to keep up with the movement of the target. If such a factor assumed the characteristics of the average of the corrections needed to keep from lagging the target on all course inputs, then a lead error may develop for the slow-rate input. Similarly, in the case of the slow-acceleration input, the application of such a correction factor on the part of the operator may account for the lack of a significant total lag-error score on Day 25.

The measure of improvement in accuracy of performance used in this study was integrated error, i.e., the error between target and hairline integrated over the time of the practice trial. A second measure of performance, the lead or lag error, was also taken. However, this latter measure need not reflect the accuracy of *S*'s performance. It is possible to track very poorly but not produce either a significant lead or lag score, i.e., by tracking in such a manner that the total amount of lead error is equivalent to the total amount of lag error. The high correlations found between integrated error and lead-lag error and the discrimination made between some conditions by the lead-lag error score which was not made by the integrated error score would indicate that the lead-lag error score is a measure of a change in behavior which underlies a change in accuracy of performance as measured by the integrated error score. The results of this experiment indicate that the improvement in accuracy of performance with practice is accomplished by a change in *S*s behavior which is analogous to a change from a one-integrator mechanism to a two-integrator mechanism.

The approach used in the present study differs from previous studies (**4, 5, 6**), which attempt to write an equation that would represent the transfer function of the human in a man-machine system. Instead of writing an equation for the operator, the present study seeks, so to speak, to present a flow chart of the analog computations which are necessary and sufficient to reproduce the *S*'s

behavior. The results of the study are relevant to the previous transfer-function studies in that it has been shown that (*a*) the human transfer function changes with learning, (*b*) the direction of the change is from that of a function which has only one integration to one which has two, and (*c*) that the limit of the change, at least under the conditions of practice in the study, is to a two-integration function.

Summary

The study attempts to make a comparison between the performance of a human operator in a continuous control system with that of the mathematically simplest mechanism which might be substituted to perform the operator's task with zero error. Two types of inputs were used, i.e., the course to be tracked by *S* was of either a constant rate or a constant acceleration. With a constant rate input, the replacement of the operator with a mechanism consisting of two cascaded integrators with appropriate feed-forward loops would reduce the error to zero. If the input is a constant acceleration, the mechanism should consist of three cascaded integrators with feed-forward loops to reduce the error to zero.

The results of the study indicate that the type of mechanism which may be substituted to provide performance analogous to that of the human operator differs as a function of the amount of practice the operator has had with the system. In general, it was found that at the beginning of practice *S* performs analogously to a one-integrator system with feed-forward loop; at the end of the amount of practice imposed by the conditions of this study *S* performs analogously to a two-integrator system.

References

1. Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. U. S. Naval Res. Lab. Rep., 1954, No. 4333.
2. Dixon, W. J., & Mood, A. M. The statistical sign test. *J. Amer. statist. Ass.*, 1946, **41**, 557–566.
3. Friedman, M. Use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
4. James, H. M., Nichols, N. B., & Phillips, R. S. (Eds.) *Theory of servo-mechanisms.* Radiation Lab. Series No. 25, New York: McGraw-Hill, 1947. Pp. 360–368.
5. Raggazini, J. R. Engineering aspects of the human being as a servo-mechanism. Paper read at Amer. Psychol. Ass. meeting, Boston, 1948.
6. Tustin, A. The nature of the operator's response in manual control, and its implication for controller design. *J. Inst. elec. Engrs*, 1947, **94**, 190–202.
7. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.

(Received April 13, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

LEARNING AND PERFORMANCE IN A COMPLEX TRACKING TASK AS A FUNCTION OF VISUAL NOISE [1]

GEORGE E. BRIGGS, PAUL M. FITTS

The Ohio State University

AND

HARRY P. BAHRICK

Ohio Wesleyan University

In most physical systems the output may contain random components that cannot be predicted exactly from knowledge of either the input signal or the known characteristics (transfer function) of the system. This uncertainty of output is labeled *noise*, where noise is used in the sense of unwanted information.

If a human operator is required to observe error in the output of such a physical system and to introduce corrections on the basis of this observation, his ability to optimize system performance is limited by the nature and amount of random variation present in the feedback channel of the man-machine system. Specifically, knowledge of results as well as identification of the stimulus may be obscured by the presence of visual noise in the information display of the system.

We would certainly expect such a disturbance to interfere with performance, but it is more difficult to predict what effect noise would have upon learning. There is some indication that the knowledge-of-results variable may have a greater effect on proficiency of performance than on long-range learning. In support of such a thesis Bilodeau (1) notes that three knowledge-of-results procedures did not produce differences in day-to-day learning effects on the SAM Multidimensional Pursuit Test, although these same procedures did give differential results at the intertrial level—momentary performance effects.

Since visual noise may obscure knowledge of results, and since noise is an inherent characteristic of most physical systems, it becomes important to determine the effects of this variable on both learning and performance. Further, it is important to determine training procedures which may produce maximum proficiency in tasks involving noise in the feedback loop of the system. The present investigation was designed to answer certain questions concerning training procedures for such tasks. It was planned as a transfer of training study wherein the transfer task contained four levels of noise amplitude in the feedback channel to the visual display. Both training and transfer tasks involved two-dimensional compensatory tracking procedures.

[1] This research was conducted in the Laboratory of Aviation Psychology of The Ohio State University and was supported in part by the United States Air Force under Contract No. AF 18(600)-1201 with the OSU Research Foundation, monitored by Dr. J. A. Adams, who served as Scientific Officer for the AF Personnel and Training Research Center. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

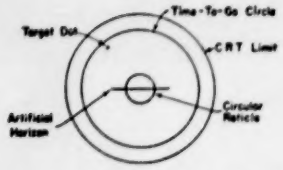

FIG. 1. The tracking display as it appeared on a 5-in. CRT.

METHOD

Apparatus.—The OSU Pilot Training Research Simulator (**3**) served as the skill task. This is an analog computer which serves as a part-task simulator of an aircraft with a radar fire-control system.[2] The *S* views a cathode ray tube (CRT) display which presents four pieces of information: (*a*) an artificial horizon, (*b*) a circular reticle, (*c*) a target dot, and (*d*) a time-to-go circle. A view of this display is presented in Fig. 1.

Before the start of each trial the target dot is given an initial displacement from the center of the reticle. Therefore, the initial system input is a step function and one may then consider the target "course" to be one of constant velocity. Disregarding for the moment the effects of the noise source, subsequent positions and movement of the target dot on the CRT are determined solely by (*a*) signals from *S*'s spring-centered control stick, and (*b*) a feedback signal, shown in Fig. 2 as 1/*T*, which was used to simulate changes in the observed target error signal associated with the approach of the attack aircraft to the constant velocity target. Lateral movements of *S*'s control generate one signal voltage, *X*, while fore and aft displacements produce another signal, *Y*. The *X* signal is subjected to two pure integrations as shown in Fig. 2. The output of the first integration drives a sine-cosine resolver which generates the angle of the artificial horizon as it appears on the CRT. The lag associated with the first integration simulates the time required for the aircraft to roll into a bank attitude. The second stage of integration moves the target dot across the face of the CRT at a rate proportional to the degree of bank. The time lag associated with this second integration simulates the rate of change of aircraft heading produced by a given bank attitude. Such integrations are accomplished by use of an EASE analog computer.[3] In the present experiment the constants of integration were the same on all trials, i.e., all trials simulated a particular airspeed, air density, etc.

The *Y* signal, from fore and aft displacements of the control stick, is transformed by a second order oscillatory function plus one integration as shown in Fig. 2. Thus, for a sudden change in fore and aft stick position the target dot and artificial horizon undergo a dampened vertical oscillation prior to assuming a new position. Again, the analog computer affected these transformations. As indicated by Fig. 2, the vertical position of the artificial horizon is affected by the output of the oscillatory exponential lag transformation while dot position is determined by the output of both the exponential and the integration stages.

The information display employed here is compensatory and the display-control relationships are of the "fly-to" variety. Thus, if the target dot is in an upper quadrant of Fig. 1, a backward deflection of the control will cause the target to drop on the CRT. If the dot is in a left quadrant of the display, a stick deflection to the left will cause the target dot to move toward the right of the CRT. In order to effect a change along the *X* coordinate, *S* first deflects the control laterally until a desired degree of

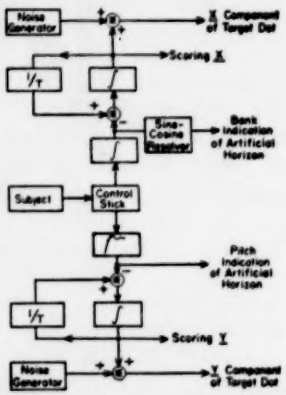

FIG. 2. A flow diagram showing major transformations of control stick signals, the points where the output voltages were scored, and the points where noise was added. The two types of transformations are an integration and an oscillatory exponential function, shown here by symbols in boxes, and a summing operation.

[2] The operational aircraft simulated, in part, was the F-86D All Weather Interceptor with its E-series fire-control system. Simulation specifications are fully described elsewhere (**3**) and are not reported here.

[3] Beckman Instruments, Inc., Berkeley Division.

TABLE 1

PHYSICAL SPECIFICATIONS, IN VOLTS, OF THE FOUR NOISE AMPLITUDE DISTRIBUTIONS
(Based on data from the Y channel of the noise source)

	Noise Level			
	N_0	N_1	N_2	N_3
RMS (volts)	0	1.25	1.88	3.17
Approximate range (volts)	0	8	12	20*

* 20 volts was equivalent to 1 in. on the 5-in CRT.

bank has been achieved; he then centers the stick while the position of the target dot changes at a rate determined by the angle of bank, and subsequently he removes the bank attitude by a deflection opposite to the original correction when it is desired to stop the lateral movement of the target dot on the CRT, i.e., when zero X error is achieved. In order to effect a change along the Y coordinate S merely displaces the control fore or aft until the target dot shows zero Y error and then he may center the stick.

A two-channel, random noise generator fed the system at the points shown in Fig. 2. When noise is added, the target dot on the CRT wanders in a random manner around the value which represents the true output of the physical system. The amplitude distribution of the disturbance used in the present study was normal or Gaussian and the power spectral density was peaked at approximately 10 cpm with 95% of the power distributed below 40 cpm. Four levels of noise amplitude were selected for this study, and Table 1 presents physical specifications for these noise levels. As shown, a noise-free condition, N_0, was included as the fourth condition. The noise signal in the X channel was statistically equivalent but uncorrelated to that in the Y channel for each amplitude level.

Tracking performance was scored prior to the introduction of the noise signals, as shown in Fig. 2. The score of primary interest here is in a time-on-target score which was cumulated for the amount of time S kept the "true" target position error within $\frac{1}{4}$ in. of the center of the target circle simultaneously in X and in Y.[4] This score was taken over the final 3 sec. of each trial, the most critical period in such tracking situations. Since the attack craft has then closed on the target, the amount of maneuvering is reduced at this time and the display-control sensitivity, which is a function of angular error in aircraft course, increases rapidly. Thus, small errors are quickly magnified and tracking becomes increasingly difficult during the final portion of the attack run.

The change in control-display sensitivity was achieved by multiplying the error signal by a factor $1/T$ as shown in Fig. 2. Since T represents time-to-go in seconds, the smaller the value of T the more sensitive the system becomes to operator error. $1/T$ was continuously variable from 1/40 sec., at the beginning of a trial, to 1 sec. at the end.

During the final 20 sec. of each trial S was given information regarding the time remaining as the time-to-go circle gradually diminished from its initial 4-in. diameter and approached the size of the circular reticle. During the initial 20 sec. the reticle had a diameter of 1 in.; during the final 20 sec. it had a diameter of $\frac{1}{4}$ in. The size of the artifical horizon remained constant throughout the trial.

Subjects.—Forty-eight junior and senior students served as *Ss*. All were volunteers from AFROTC classes and each S was paid for five experimental sessions of approximately 45 min. each. They served in pairs; the first S received five trials at the task while his companion observed, then the second S received five trials at the task and so on until each had received a total of 20 trials. There were five such sessions or a total of 100 trials for each S. Scoring was cumulated for five-trial blocks.

Procedure.—During the first session E gave each pair of *Ss* a demonstration of the task requirements. Then a total of 20 40-sec. trials were given to each S with 20-sec. rest periods between trials. After each block of five trials S received a 5-min. rest period during which his partner practiced the same task. A 24-hr. interval separated all sessions.

TABLE 2

SUMMARY OF THE EXPERIMENTAL TREATMENTS

Group	Sessions				
	1	2	3	4	5
1	N_0	N_0	N_0	N_0	N_x*
2	N_0	N_1	N_1	N_1	N_x
3	N_0	N_2	N_2	N_2	N_x
4	N_0	N_x	N_x	N_x	N_x

* The N_x or mixed noise condition consisted of blocks of five trials at each of the four noise levels N_0, N_1, N_2, and N_3.

[4] Generation of the target signal and the scoring of error was done in rectangular coordinates. The time-on-target scoring area was a square $\frac{1}{4}$ by $\frac{1}{4}$ in. in size.

Table 2 presents a summary of the daily treatments of the four groups during the five experimental sessions. These sessions covered (*a*) pretraining—Session 1, (*b*) training—Sessions 2 through 4, and (*c*) transfer—Session 5. There were 12 *S*s randomly assigned to each of four groups. During Session 1 all *S*s, regardless of group, received 20 trials under the no-noise, N_0, condition. This permitted standardization of instructions for all groups and gave *E* a means of checking for comparability of group performance prior to the training period.

The *S*s of Group 1 continued to practice at the simulator task under Condition N_0 for three additional sessions while Groups 2 and 3 performed under noise conditions N_1 and N_2, respectively, for these 60 training trials. Group 4 received five trials on each of the four noise conditions N_0, N_1, N_2, and N_3 during each of the three training sessions. A randomized 4×4 latin square was used to balance sequence effects for Group 4, the letters of the square representing noise conditions, the rows the *S*s, and the columns five-trial blocks. The same square was used for each of the three training sessions. During Session 5 all *S*s transferred to this varied noise treatment; each *S* was given five consecutive trials on each of the four noise conditions. The same latin square used for Group 4 on Sessions 2, 3, and 4 served for all groups on the transfer trials on Session 5.

The instructions to *S* emphasized that the "true" position of the target dot was being obscured by the visual noise. He was instructed, then, to "fly to the average" of the random motion as this was the best estimate of the target signal. In other words, *S* was instructed to attempt to smooth out or disregard the random component of target motion and track through the noise.

Results

Figure 3 presents graphically the training and transfer data for all groups over the five sessions of the experiment. Each point is the mean performance of 12 *S*s as determined by the time-on-target scores averaged for five-trial blocks.

Analyses of variance for correlated measures were made for the following data: (*a*) all groups during Session 1, (*b*) Groups 1, 2, and 3 for Sessions 2, 3, and 4, (*c*) Group 4 for Sessions 2, 3, and 4, and (*d*) all groups for Session 5, the transfer trials. Prior to analysis all scores were transformed to arc sin (percent time on target)$^{\frac{1}{2}}$ (5).

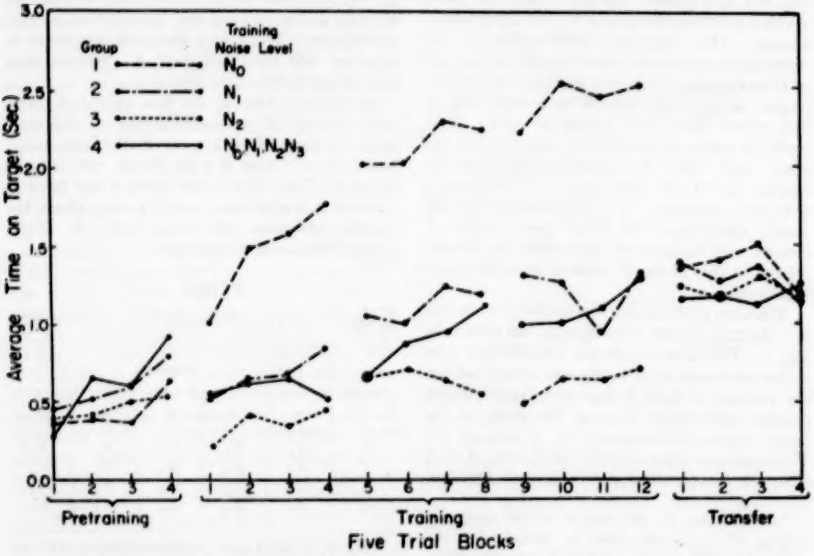

Fig. 3. Training and transfer performance levels for all groups over the five experimental days. See text for description of transfer session. The maximum score was 3 sec.

This transformation was necessary here as both a lower (0 sec.) and an upper (3 sec.) limit occurred in these distributions of scores.

The results of the first analysis are as follows: During pretraining (Session 1), when all groups received the N_0 condition, no statistically significant differences were obtained for the between-groups comparisons. There was, however, a significant ($P < .01$) increase over the four blocks of trials in the proficiency of tracking for all groups combined. Therefore, the data of Session 1, Fig. 3, are interpreted as indicating (*a*) an increase in tracking proficiency from an initial average level of .4 sec. on target to an average of approximately .7 sec. (out of a possible 3 sec.), and (*b*) no reliable differences between groups prior to the training trials.

Table 3 summarizes the analysis for Groups 1, 2, and 3 during training (Sessions 2, 3, and 4). A separate analysis for Group 4 over the same interval is listed also. From the three-group analysis the comparisons that are of primary interest are those for Groups, Sessions, and the Groups × Sessions interaction. All are statistically significant at or beyond the $P < .05$ level. The significant interaction justified the use of the Duncan Multiple Comparisons Test (**2**) for an analysis of the pattern of differences among the nine means which define that interaction. Sessions × Blocks × Ss within groups served as the error estimate for this test, and the $P < .05$ level was accepted for significance.

The results of this test are illustrated in Fig. 3. The proficiency of Group 1 increases until at least the second block of trials on Session 4 whereas Group 2 approaches an asymptote by the third block of Session 3, and Group 3 shows no further increase in proficiency from that attained during the first two blocks of Session 3. Thus, not only do increased amplitudes of noise produce lower levels of tracking proficiency but it appears that the greater the noise amplitude the earlier in training does performance approach an asymptote. All groups had approached an asymptote at or before the time of transfer.

The companion analysis listed in Table 3 for Group 4 shows that a statistically significant increase in performance occurs over Sessions 2,

TABLE 3

ANALYSES OF VARIANCE FOR THE DATA OF SESSIONS 2, 3, AND 4

Source	*df*	Mean Square	*F*
Groups 1, 2, and 3			
Groups (*G*)	2	43,960.90	74.77*
Ss/Group (*Ss*/*G*)	33	589.93	
Sessions (*D*)	2	7,623.50	48.70*
D × *G*	4	748.76	4.78*
D × *Ss*/*G*	66	156.54	
Blocks (*B*)	3	483.70	3.95*
B × *G*	6	237.77	1.94
B × *Ss*/*G*	99	122.60	
B × *D*	6	156.66	1.78
B × *G* × *D*	12	106.60	1.21
B × *D* × *Ss*/*G*	198	88.23	
Group 4			
Order (*O*)	3	160.44	—
Ss/Order (*Ss*/*O*)	8	1,369.00	
D	2	2,546.13	13.35*
O × *D*	6	15.76	—
D × *Ss*/*O*	16	190.59	
Noise level (*N*)	3	5,288.99	31.31*
N × *D*	6	292.88	1.73
B	3	329.83	1.95
B × *D*	6	109.42	—
Square uniqueness (*U*)	6	147.86	—
U × *D*	12	94.56	—
Residual error	72	168.91	

* Significant at or beyond the $P < .05$ level.

3, and 4, as was the case with Groups 1, 2, and 3. In addition to this, visual noise level appears as a significant source of variance. This latter finding agrees with results of the three-group analysis where noise level defines the group variable.

Analysis of the data for the transfer trials, Session 5, showed no significant differences among the four groups. Further, there was no significant increase or decrease in the level of performance over the four blocks of trials of that session. The variable of noise level was again statistically significant ($P < .01$) as would be expected from the analyses of the data from the training sessions.

Despite the lack of any significant differences among groups, Fig. 3 shows that the session means for each of the three experimental groups exceed that of the control group (Group 4) during the transfer session. Therefore, all three conditions of specific training produced 100% or greater transfer as compared to Group 4 performance. This might have been predicted for Groups 1 and 2 from the higher levels of performance attained by these two groups during the training trials, but would not have been predicted for *S*s of Group 3, who appeared to be quite inept at the task on the training trials. Nevertheless, the proficiency of this latter group is at least equal to that of the control group on the transfer trials.

It should be recalled that the four points in Fig. 3, which define each of the transfer curves, are the average performance of groups over all four noise levels, i.e., these points are the column means of the 4×4 latin square for each group. The comparability of groups suggested by Fig. 3 is shown even more clearly in Fig. 4, where the performance of

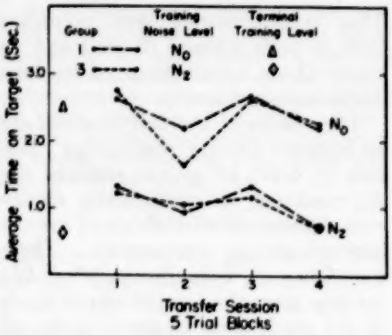

FIG. 4. A comparison of Groups 1 and 3 on transfer noise levels N_0 and N_2. The maximum score was 3 sec.

Groups 1 and 3 are presented for transfer noise levels N_0 and N_2. This shows that even though Group 1 had never previously experienced N_2, they were of equal proficiency to Group 3 who had practiced for 60 trials on that particular noise level. It is even more remarkable that Group 3 was comparable to Group 1 on N_0, the noise-free tracking condition, even on the first block of transfer trials. Again, this would seem to be incongruous with the level of performance attained by Group 3 during the training sessions Further, Group 1 required 60 trials to attain that level of proficiency on N_0 and it appears that Group 3 had attained a comparable level almost immediately. An hypothesis, which could account for these results, is discussed in the following section.

DISCUSSION

It is obvious from Fig. 3 that visual noise markedly affected performance in this compensatory tracking situation: increases in the uncertainty of information feedback were accompanied by decreases in the level of tracking performance during the training trials of Sessions 2, 3, and 4. However, on the

transfer trials (Session 5), during which all groups received five consecutive trials on each of four noise conditions, all groups were comparable in their tracking proficiency. Therefore, we may conclude that changes in the reliability of information feedback can reduce *performance*, and yet it apparently has little net effect upon *learning* in a situation such as that studied. This, of course, agrees with the earlier findings of Bilodeau (**1**). Our concern here will be to propose an hypothesis to account for the differential effects of visual noise on performance and on learning in this particular type of task.

Performance effects.—The primary effect of visual noise was to obscure the "true" location of the target signal as it appeared on the tracking display. This effect required that *S* attempt to discriminate signal ("true" location) from noise (the random component of target location). This discrimination is possible only when *S* bases his response on a time sample of the stimulus, i.e., he must perform an operation on the stimulus analogous to an integrator in order to estimate the mean of the two-dimensional disturbance. Such discrimination becomes increasingly difficult as the amplitude of the random disturbance is increased, for the greater the level of noise, the longer the time sample which *S* must take before he is able to discriminate the true error component with equal accuracy as distinct from the random component of the target motion. This increase in difficulty of discrimination between signal and noise could easily account for the differential levels of performance under the different noise levels during the training sessions.

Training effects.—Both incidental observation and examination of graphic records of stick deflection reveal two ranges of corrections typical in this task. The first is a large amplitude adjustment which is achieved by a relatively large stick deflection of considerable duration. These large amplitude responses move the target dot with considerable acceleration and this affects a gross change in the position of that signal on the tracking display. These corrections are typical when a practiced *S* perceives a relatively large error, as at the beginning of a trial.

The second range of adjustments are relatively small amplitude corrections which serve either to reduce small errors or to compensate for residual target drift as the sensitivity of the system increases toward the end of each trial. These fine corrections are analogous to the "current control" movements first described by Woodworth (**8**). The term *primary movement* has been coined to identify the initial, large amplitude adjustments, and "current control" or *secondary movements* are small amplitude discontinuities which serve to complement the error reduction achieved by the initial, gross adjustment. Several studies have been concerned with the parameters defining the "current control" type of response in simple corrective movement tasks, and it has been accepted as a general characteristic of tracking performance (**4**, **6**, **7**).

On the one hand, it is apparent that *S* can discriminate between the effects of visual noise and the effects of his own primary, large amplitude corrections since in this case target rates and position changes are considerably greater in amplitude than those associated with the random fluctuation due to noise. On the other hand, small amplitude corrections produce effects whose amplitudes are well within the range of the random disturbance of the target signal. Therefore, when noise is present *S* receives relatively less reliable information concerning the effects of his secondary adjustments. For Groups 2 and 3, then, one would expect knowledge of results to be restricted to primary adjustments, the effects of small amplitude corrections being masked by the random disturbance called visual noise. Therefore, we assume that the effects of primary adjustments were learned by both the noise-free group and those groups trained with visual noise on the target signal. However, we also assume that only the noise-free situation, Group 1, permitted significant learning of the effects of secondary, small amplitude

corrections. Some *S*s who tracked in the presence of visual noise may have made many small corrective movements, of course; it is the absence of knowledge of results, not the absence of responses, that we are emphasizing.

At first glance, it would appear that maximum proficiency in the present tracking task would demand practice with knowledge of results of both the primary and secondary adjustments. However, if the above line of reasoning is accepted, this is apparently not the case as the training schedule of Group 3 placed major restrictions on any learning specific to the effects of secondary corrections, and yet, during the transfer session, this group did not differ significantly from Group 1 on any of the noise levels including the noise-free trials (see Fig. 4).

We, therefore, propose the following hypothesis to account for the negligible effects of visual noise on the development of skill in this tracking situation: the attainment of proficiency in a compensatory tracking task is dependent upon the opportunity to learn the skillful execution of primary, large amplitude corrections and (if necessary) secondary, small amplitude corrections. However, if practice on secondary adjustments is limited, as is assumed to be the case with noise in the feedback channel, then *overlearning* of only the primary corrections is sufficient preparation for maximum proficiency in a subsequent task requiring both primary and secondary corrections. This hypothesis is limited by the following restrictions: (*a*) the primary and secondary adjustments may differ by degree but not in kind, and (*b*) the training task may restrict knowledge of results from secondary corrections but must preserve the major interactions and transformations of the transfer system.

The first restriction implies that perceptual-motor learning is not specific to any arbitrarily designated range of response amplitudes but is expected to generalize readily over the major extent of such range, especially with overlearning. However, learning is not expected to generalize so readily to other types of adjustment; corrections practiced during the training task which are qualitatively different from those required during transfer would not necessarily support the hypothesis. The second restriction implies that for maximum proficiency during transfer, the training task should represent a *simplified* model of the total transfer system and not a task composed of an isolated element of the transfer task such as has been used in many part-whole learning situations.

It should be acknowledged that this primary adjustment hypothesis merely elaborates an observation by Woodworth in his classic study on simple corrective movements (**8**). Although our experimental task was considerably more complex than that utilized by Woodworth, our hypothesis is in agreement with his statement that "the path to skill lies in increasing the accuracy of the initial adjustment" (**8**, p. 59).

Summary

The present report describes a transfer of training study wherein the transfer task contained four amplitude distributions of low frequency visual noise in the feedback channel of a complex man-machine system. The task simulated an aircraft with a radar fire-control system wherein *S* was required to perform under compensatory tracking conditions. Four groups of 12 *S*s each served for five days and each *S* received 20 40-sec. trials per day.

On the first session all *S*s practiced under a no-noise, N_0, condition. On Sessions 2 through 4 Group 1 continued tracking with N_0, Group 2 performed under noise disturbance N_1 (a low level of noise amplitude), Group 3 tracked under noise condition N_2, (a larger amplitude of noise), and Group 4, the control group, received varied training (five consecutive trials on each of four noise conditions, N_0, N_1, N_2, and N_3) during each session. On the fifth session, the three experimental groups transferred to the same varied noise condition.

On Session 1, all groups achieved comparable levels of proficiency at the task. Over the three training sessions, the N_0 group attained the highest level with terminal performance at 81% of maximum proficiency. The terminal performance of the N_1 and N_2 groups was 40% and 20% of maximum, respectively, while the group

trained under the varied noise condition achieved an average terminal performance level of 37% of maximum. Asymptotic levels were approached by all groups at or before the transfer trials.

On the transfer trials all groups were found to be of comparable proficiency on the mixed noise condition; therefore, it is concluded that the several training procedures did not differentially affect the skill required to track under varied conditions of noise. From these results one may conclude that whereas visual noise may depress performance levels, it apparently has little net effect upon learning in this skill task. A possible explanation for this differential effect of noise was proposed as an extension of Woodworth's two-component theory of corrective movements (**8**).

References

1. Bilodeau, E. A. Variations in knowledge of component performance and its effects upon part-part and part-whole relations. *J. exp. Psychol.*, 1955, **50**, 215–224.
2. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2**, 171–189.
3. Harter, G. A., & Fitts, P. M. The functional simulation of complex systems by means of an analog computer, with the F-86D, E-4 system as a specific example. AF Personnel and Training Research Center, *Res. Rep. AFPTRC-TN-56-133, Part I*, December, 1956.
4. Hick, W. E. The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1**, 36–51.
5. Snedecor, G. W. *Statistical methods*. Ames, Iowa: Iowa State College Press, 1948.
6. Vince, M. A. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, **38**, 149–157.
7. Welford, A. T. The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, 1952, **43**, 2–19.
8. Woodworth, R. S. The accuracy of voluntary movement. *Psychol. Monogr.*, 1899, **3**, No. 2 (Whole No. 13).

(Received May 18, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

PERCEPTION OF NONSENSE PASSAGES IN RELATION TO AMOUNT OF INFORMATION AND SPEECH-TO-NOISE RATIO[1]

IRWIN MILLER[2]

Purdue University

Attneave has recently pointed out an important correspondence between information distribution and the Gestalt principles of perceptual organization that leads him to interpret the *good Gestalt* as a figure with high internal redundancy (**2**). In agreement with this interpretation, an auditory experiment reported by Cherry (**3**) showed that in simultaneously spoken messages highly redundant verbal groups (clichés) are perceived correctly, even though the messages are otherwise confused. Two learning studies that also bear directly on what is conceptualized here as verbal redundancy are Thorndike's classic experiment on "belongingness" (**9**) and Guilford's demonstration of the role played by "meaningfulness" (**5**).

The present research was designed to investigate functional relationships between auditory perception and two independently varied speech attributes: the amount of information expressed in the patterning of the speech ("cliché"-edness, "belongingness") and the intensity of the speech relative to white noise. The techniques employed have been adapted from a study by Miller, Heise, and Lichten (**7**) dealing with relationships between speech intelligibility and context, and a study by Aborn and Rubenstein (**1**) in which information theory was applied to a verbal learning situation.

Miller, Heise, and Lichten (**7**) determined thresholds for digits, words in sentences, and nonsense syllables presented through headphones at various speech-to-noise ratios (S/N's). With two of the researchers alternating as *E* and *S*, they found 50% thresholds at S/N's of − 14 db (digits), − 4 db (words in sentences), and + 3 db (nonsense syllables). These differences were interpreted as evidence that intelligibility depended on the range of possible alternatives associated with each type of material.

Aborn and Rubenstein (**1**) constructed passages of nonsense syllables according to rules based on the logarithmic definition of information (*H*) derived by Shannon (**8**). The organizational structure of the passages was specified exactly, in bits, and treated as an experimental variable. They found that the mean number of correctly recalled syllables tended to increase monotonically from the 4-bit passage to the 1-bit passage, and that the amount of information gained, computed as the product of the number of syllables recalled and average amount of stimulus information, tended to remain constant for most of the passages.

METHOD

The *S*s were first made familiar with the structure and classification scheme of a nonsense

[1] This paper is based on a dissertation submitted in 1954 to the Graduate School of Purdue University in partial fulfillment of the requirements for the Ph.D. degree. I am indebted to Dr. Lawrence M. Baker for his encouragement and guidance throughout the course of the research. I would also like to acknowledge gratefully the contribution of the Voice Communication Laboratory of the Purdue Speech Department in providing skills and facilities essential to the research.

[2] Now with Bell Telephone Laboratories at Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

vocabulary through combined visual and auditory training. Following this, *Ss* listened to different tape-recorded nonsense passages, writing down what they heard, while white noise at various S/N's was also admitted into the headphones.

Several preliminary steps included constructing nonsense passages intended to represent different amounts of input information, hereafter referred to as $H(x)$, making tape recordings of the passages and the nonsense vocabulary they were based on, and assembling an instrumental circuit for the presentation of speech with noise.

Preparation of verbal materials.—A vocabulary of 16 nonsense dissyllables was constructed by systematically permutating four consonants about two vowels. This was done to make the consonants the prime information carriers and to distribute transitional cues. In order to construct different passages according to simple organizational rules, the 16 items were divided into four classes as shown in Table 1. The members of Class 1 all begin with M, Class 2 items all begin with R, Class 3 with V, and Class 4 with Z. In addition, each class is subdivided into two pairs of items: those having a medial V and those with a medial Z.

Sixteen passages were then constructed, each 36 items long. Four of the passages were organized to represent $H(x)$ at the rate of 1 bit per item, four at the rate of 2 bits per item, four at 3 bits, and four at 4 bits. The rules by which the passages were formed allowed all of the items equal probability within each passage.

In the four 4-bit passages, the dissyllables were simply selected at random from the vocabulary.

In the four 3-bit passages, the successive items were selected alternatingly from the V subclass (1V, 2V, 3V, 4V) and the Z subclass (1Z, 2Z, 3Z, 4Z). This restriction excluded the possibility of two successive items having the same medial consonant, reducing the number of alternatives at each point to eight.

In the four 2-bit passages, the items were arranged in groups of four. Within each group, the first item was selected from Class 1, the second from Class 2, the third from Class 3, and the fourth from Class 4. This pattern reduced the number of alternatives at each point to four, throughout the length of the passage.

In the four 1-bit passages, the items were arranged in groups of four, as in the 2-bit passages. As an additional restriction, selection in the successive groups of four was limited to the V subclass for the first group, the Z subclass for the second group, the V subclass for the third group, and so on. This pattern reduced the alternatives at each point to two.

Recording and instrumentation.—The nonsense vocabulary was recorded by means of an Altec-Lansing Model 11 Microphone System with an Altec 21-B miniature Condenser Microphone and a Stancil-Hoffman Magnetic Tape Recorder, Model R-4. The talker, a professor of voice science, monitored his speaking level with the volume indicator (VU meter) of the tape recorder, maintaining the peak intensities within ± 1 db of the recording level. He made efforts to minimize differences in stress and duration among the 16 items.

The items of this master tape were then dubbed repeatedly to prepare 16 passages by directing the output of the Stancil-Hoffman recorder into a Presto Tape Recorder, Type 900A. An additional tape, featuring two consecutive presentations of the vocabulary list, was also dubbed for use in training the *Ss*. The *E* employed the Altec-Lansing microphone system to record supplementary material on this training tape in his own voice, namely, the number and letter preceding each pair of nonsense items, e.g., "1 V . . . meevoom . . . meevoor."

The nonsense passages were combined with white noise by passing the output of a General Radio Random-Noise Generator, Type No. 1390-A, through a Hewlett-Packard Attenuator Set, Model 350A, and combining it with the output of the Presto tape recorder in a Thordarson Amplifier, Model T-31W25A. The combined speech and noise were then directed to a pair of Utah-Chicago Receivers, Model ANB-H-1, which *S* wore. The S/N was varied by changing the level of the white noise.

TABLE 1

NONSENSE VOCABULARY

Class	Subclass	Dissyllables
1	V	MEEVOOM
		MEEVOOR
	Z	MEEZOOV
		MEEZOOZ
2	V	REEVOOZ
		REEVOOM
	Z	REEZOOR
		REEZOOV
3	V	VEEVOOV
		VEEVOOZ
	Z	VEEZOOM
		VEEZOOR
4	V	ZEEVOOR
		ZEEVOOV
	Z	ZEEZOOZ
		ZEEZOOM

Before each *S* took part in the perception experiment, the voltage across the terminals of the headphones was measured independently for the tape recorder and the noise generator with a vacuum tube voltmeter. The tape recorder measurement was based on a 1000-cycle recorded tone. The output levels of the two instruments were then equated (S/N = 0) at a voltage equivalent to 78 db re .0002 dynes/cm.^{2} as determined by an ADC Artificial Ear. During the familiarization training, the training tape was presented by means of a Knight Tape Recorder, Model 96-485.

Subjects.—Thirteen *S*s took part in this study; 10 were graduate students, 2 were undergraduates, 1 had no college experience. The *S*s ranged in age from 22 to 32, the modal age being 25. There were nine males and four females.

Familiarization training.—A procedure combining visual and auditory presentation was used to familiarize *S*s with the set of nonsense items and the classification scheme. The *S* was seated so that he could easily look at a standard memory drum by glancing to one side. The *E* rotated the memory drum manually, presenting each pair of nonsense items visually at the same time they were heard from the loudspeaker of the tape recorder.

On the introductory trial *S* listened to and copied each pair of nonsense items as it was presented by the training tape and the memory drum preceded by a letter-number combination. After the first trial *S* was told to study the array of nonsense items for 2 min. Drill trials were then begun with *S* instructed to anticipate or guess the pair when each of the associated number-letter combinations was heard. When *S* made an error he was told to copy both members of the pair from the memory drum; when *S* was correct, the appropriate spaces on the answer sheet were left blank. On the training tape the spoken number-letter preceded the associated spoken nonsense items by about 8 sec. From the end of a pair to the onset of the number-letter of the next pair, there was an interval of about 9 sec. The memory drum was rotated so that

1 V MVM MVR

appeared when the recorded voice began to say "meevoom." The abbreviations shown above (i.e., MVM = meevoom) were used throughout the experiment to simplify *S*'s task.

The successive presentations were made in a regular order that was consistent for each *S*, but was varied among *S*s by advancing the starting point to distribute proactive and retroactive inhibition. The training was continued with 2-min. rests between trials until each *S* attained the criterion of two consecutive perfect trials.

Perceptual experiment.—The four nonsense passages that were organized to represent $H(x)$ at the rate of 1 bit per item were randomly assigned to S/N values of − 13 db, − 6 db, + 1 db, and + 8 db. The four passages representing $H(x)$ at the rate of 2 bits per items were similarly assigned, as were the four passages representing 3 bits per item, and the four representing 4 bits per item. A random presentation order for the 16 conditions was made up for each *S*.

The *S*s were told to write down the nonsense items as they heard them, guessing when they were not sure. Before each passage was presented, *E* gave *S* a card bearing an appropriate organizational rule. The *E* answered questions raised by *S*, making sure that the rules were understood.

The interval between items was about 9 sec. in all passages and *S* recorded his response during this interval. An effort was made to limit *S*'s responses to the reproduction of actual members of the nonsense vocabulary by providing a list of the 16 items for optional reference during the experiment.

Results

Familiarization training.—Although *S*s formed a fairly homogeneous group in terms of age and education, large individual differences in learning speed occurred. The number of trials required by the 13 *S*s, ranked from the fastest *S* to the slowest *S*, was: 4, 5, 5, 6, 6, 7, 7, 8, 10, 11, 13, 14, 15. When the number of trials required was correlated with total number of correct responses in the perceptual experiment, a nonsignificant r of − .36 resulted.

Perceptual experiment.—Sixteen scores were obtained for each *S*, representing the total of correctly reproduced nonsense items for each combination of S/N with $H(x)$. An analysis of variance was made to test the significance of differences associated with the experimental variables. This analysis is summarized in Table 2.

The differences associated with the four $H(x)$ rates proved significant at

TABLE 2

ANALYSIS OF VARIANCE OF PERCEPTION SCORES

Source	*df*	*MS*	*F*
$H(x)$	3	1341.782	145.97**
Linear component	1	3853.850	
Deviation	2	85.748	9.33**
Quadratic component	1	13.000	
Deviation	1	158.496	17.24**
$H(x) \times Ss$	36	9.192	
S/N	3	2784.167	168.11**
Linear component	1	8209.696	
Deviation	2	71.402	4.31*
Quadratic component	1	20.942	
Deviation	1	121.862	7.36*
S/N × *Ss*	36	16.562	
S/N × $H(x)$	9	175.718	20.85**
S/N × $H(x)$ × *Ss*	108	8.429	
Ss	12	47.277	
Total	207		

* $P = .05$.
** $P = .01$ or less.

the .001 point. The relationship between auditory perception and $H(x)$ is clearly shown in Fig. 1: a progressive decrease in the percentage of items correctly perceived results as $H(x)$ is increased. Following the method described by Cochran and Cox (**4**, pp. 55–59), a further analysis of the relationship was made by testing the residual variance after the removal of linear and quadratic components. The variance remaining after the removal of the quadratic component was significant at the .001 point, indicating that the function relating auditory perception to $H(x)$ is at least a third-order curve. With reference to Fig. 1, this analysis implies that the changes in acceleration at 2 bits and 3 bits are real features of the function.

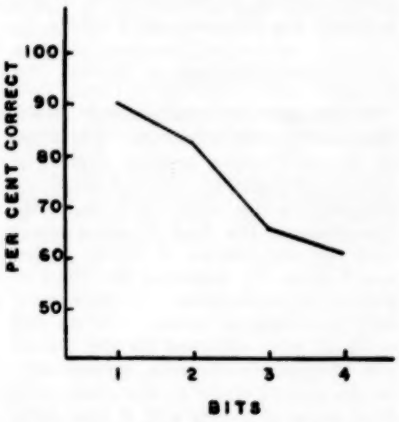

FIG. 1. Perception of nonsense dissyllables as a function of $H(x)$ in bits per item.

In order to estimate how much information was *transmitted*, calculations were made according to Shannon's equation for the discrete channel with noise (**8**, p. 407):

$$R = H(x) - H_y(x),$$

where R = rate of transmission, $H(x)$ = rate of input information, and $H_y(x)$ = rate of input information when the output is known (equivocation). A summary of the R rates as a function of $H(x)$ and S/N is given in Table 3.

The differences associated with the four S/N's were significant at the .001 point. This relationship with auditory perception is shown in Fig. 2. A progressive increase in the percentage of items correctly perceived results as the S/N is increased. Since the four S/N's were spaced at equal

TABLE 3

TRANSMITTED INFORMATION, R, AS A FUNCTION OF INPUT INFORMATION, $H(x)$, AND SPEECH-TO-NOISE RATIO, S/N

$H(x)$	S/N				All S/Ns
	+8 db	+1 db	−6 db	−13 db	
1	.86	.71	.42	.30	.53
2	1.92	1.60	1.20	1.04	1.32
3	2.63	2.34	2.00	2.08	2.07
4	3.63	3.16	3.00	3.14	3.04

Note—Information measures are in bits per dissyllable.

steps of 7 db, a further analysis of the variance was possible, comparable to the analysis of the $H(x)$ function. The variance remaining after the quadratic component was removed proved significant at the .05 point, indicating that the S/N function is also a third-order curve, at least. This analysis implies that the changes in acceleration are real characteristics of the S/N function.

The interaction of S/N × $H(x)$ proved significant at the .001 point, indicating that the differences in percentage of items correct associated with the various $H(x)$ rates depend on the particular S/N at which the differences are observed. The manner in which the two variables interact is shown in Fig. 3. At a S/N of + 8 db, the percentage of items correct is very high at all $H(x)$ rates, showing an over-all decrease of only 4.5%. At a S/N of + 1 db, the percentage of items correct decreases regularly at every increase of input information, resulting in an over-all decrease of 21.6%. At a S/N of − 6 db, the percentage of items correct decreases 40.6% as the information is increased. Finally, at a S/N of − 13 db, the greatest over-all decrease occurs, a difference of 53.0% between the 1-bit and 4-bit rates.

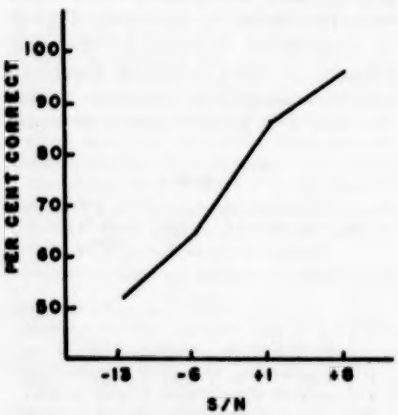

FIG. 2. Perception of nonsense dissyllables as a function of S/N. The average speech level was held constant at about 78 db re .0002 dyne/cm.².

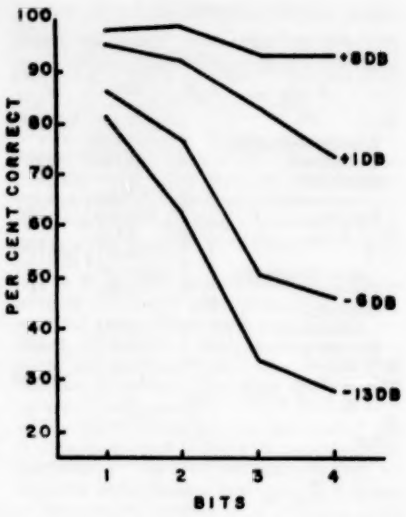

FIG. 3. Perception of nonsense dissyllables as a function of $H(x)$ in bits per item with S/N as the parameter. The average speech level was held constant at about 78 db re .0002 dyne/cm.².

DISCUSSION

In general the experimental results agree with previous findings. The shape of the S/N curve appears typical of speech intelligibility research involving wide-band white noise (cf. **6**, p. 1050). The shape of the $H(x)$ function agrees with the observations of Miller, Heise, and Lichten (**7**) regarding the effect of context on intelligibility. In their study 50% thresholds of about − 13 db and − 12 db were observed for the 8-word and 16-word vocabularies, respectively. In the present study at the comparable $H(x)$ rates of 3 bits and 4 bits, 50% thresholds resulted at about − 6 db and − 5 db, respectively. Despite these

differences in absolute thresholds, which may reflect differences in experience of *S*s, or differences in type of materials and their organization, the over-all dependence of perception on number of alternatives is substantiated.

The demonstration of the S/N × $H(x)$ interaction is of considerable interest. This interaction emphasizes the fact that difference in percentage of items perceived at the four $H(x)$ rates cannot be interpreted as a simple function of the a priori probabilities, but on the contrary, depends significantly on S/N.

According to the computed values of R, the transmitted information increases as $H(x)$ increases, notwithstanding the decreasing percentage of items correctly perceived. This relationship holds at each of the four S/N's investigated. A regular decrease of the R values over all $H(x)$ levels occurred with every decrease in S/N except for the two highest $H(x)$ levels at S/N equal to − 13 db.

Summary

This study related perception of nonsense passages to two variables: the average rate of input information, $H(x)$; and the speech-to-noise ratio, S/N. A nonsense vocabulary was taught to 13 *S*s by means of repeated auditory and visual presentations. At the same time the *S*s also learned a scheme for classifying the nonsense items. In the perceptual experiment that followed, the nonsense items were presented auditorily together with white noise. The items were arranged in passages representing four $H(x)$ rates: 1 bit (per item), 2 bits, 3 bits, and 4 bits. At each $H(x)$ rate, four passages were presented, at S/N's of − 13 db, − 6 db, + 1 db, and + 8 db.

The mean number of items correctly reproduced at the four $H(x)$ rates differed significantly ($P < .001$) with the means varying as an inverse function of $H(x)$. Tests of the residual variance indicated that the relationship was that of a third-order function or higher. The mean number correct at the four S/N rates also differed significantly ($P < .001$), with means varying as a direct function of S/N. This was also at least a third-order function.

An interaction between $H(x)$ and S/N ($P < .001$) acted to qualify their independent effects. Differences associated with $H(x)$ were slight when S/N was high; however, as S/N decreased, the differences associated with $H(x)$ increased significantly in a regular manner. Estimates of R—transmitted information—indicated that as $H(x)$ increased, R also increased, although the percentage of items correct decreased. This relationship was found at all four S/N's.

References

1. Aborn, M., & Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.
2. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.
3. Cherry, E. C. Some experiments on the recognition of speech, with one and two ears. *J. acoust. Soc. Amer.*, 1953, **25**, 539–547.
4. Cochran, W. G., & Cox, G. M. *Experimental designs*. New York: Wiley, 1950.
5. Guilford, J. P. *Laboratory studies in psychology*. New York: Holt, 1934.
6. Licklider, J. C. R., & Miller, G. A. The perception of speech. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
7. Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1951, **41**, 329–335.
8. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423; 623–656.
9. Thorndike, E. L. *Human learning*. New York: Century, 1931.

(Received June 15, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

REST AND WARM-UP IN BILATERAL TRANSFER ON A PURSUIT ROTOR TASK [1]

L. C. WALKER,[2] C. B. DE SOTO,[3] AND M. W. SHELLY [4]

University of Wisconsin

Pursuit rotor learning provides possibilities for the study of a special kind of reminiscence—bilateral reminiscence—which has been reported in a number of recent papers. Representative studies are those of Ammons and Ammons (**4**), Grice and Reynolds (**6**), Irion and Gustafson (**9**), Kimble (**10**), and Rockway (**13**). In pursuit rotor studies of bilateral reminiscence, *Ss* perform with one hand prior to rest and change to the other hand—the transfer hand—following rest. The bilateral reminiscence phenomenon has been studied principally with a view to localizing reactive inhibition (**6, 9, 10**). Results from the investigations of such presumed reactive inhibition effects suggest that these effects are not specific to the effectors exercised prior to rest but may involve more central processes. Because the different experiments have employed techniques which vary in such respects as prerest and postrest hand, trial duration, intervals between trials, duration of practice, and sex of *Ss*, the results are not strictly comparable. However, it does appear that bilateral reminiscence is similar, as a distribution of practice phenomenon, to the earlier discovered unilateral reminiscence, although, as shown by Grice and Reynolds (**6**), somewhat less in magnitude.

Introduction of rest underlies another factor in motor learning besides reactive inhibition, namely, "loss of set" or the "need for warm-up" as discussed by Ammons (**2, 3**) and Irion (**7, 8**). In a unilateral transfer situation using the pursuit rotor, Denny, Frisbey and Weaver (**5**) have shown that warm-up following 5 min. of rest is complete after 30 sec. of postrest practice, while reactive inhibition does not reach a maximum for 2.5 min. of practice. Silver (**14**) has shown on a unilateral alphabet printing task that the maximum period of postrest warm-up effectiveness for a 2.5-min. period of prerest practice is 30 sec. These results suggest that if the final 30-sec. pretransfer hand trial occurs immediately following rest and before bilateral transfer, it should serve as warm-up and improve initial performance on the transfer hand. It might be expected that this improvement, in comparison to a control group with rest but no warm-up, would disappear within a few trials as the latter group "re-establishes set" or "warms up" following resumption of practice. It thus seems feasible to test the question: Does warm-up, as well as reminiscence, occur in bilateral transfer?

METHOD

A Koerth-type pursuit rotor with a 12-in. diameter disc rotating at 60 rpm and with a ¾-in. diameter contact point or "target" was the device used. Connected to the rotor were two electric timers calibrated to read time on

[1] A modified form of this paper was presented by the first author at the Midwestern Psychological Association Convention in St. Louis, Mo., May, 1956.

[2] Now at the Officer Education Branch, Personnel Training Laboratory, AFPTRC, Maxwell Air Force Base, Alabama.

[3] Now at The Johns Hopkins University.

[4] Now at The Laboratory of Aviation Psychology, The Ohio State University.

target in 1/100 sec. The timers were wired in such a way that they alternated in recording the time on target every 30 sec., the length of a trial. All of the equipment was placed on a table 3 ft. high and the timers were situated so that *S* could not see them.

The experimental design involved the use of five conditions with rest periods, number of trials, and warm-up arrangements as shown in Table 1.

It will be noticed that *S*s in all conditions received the same amount (eight trials) of pretransfer practice, the only difference being whether the eighth trial was before or after the rest interval.

The general procedure was as follows: Each *S* was brought in singly, shown the apparatus and given a demonstration by *E* (with the rotor turned off) as to how to hold the stylus. He was told to use a free and easy swinging motion and that he would be scored for the amount of time spent on target. Before starting, each *S* was given instructions which varied slightly according to the condition to which he was assigned. The *S*s in Cond. A, C and E were given 5 sec., timed by a stopwatch, to shift from right to left hands. The *S*s with rest intervals were given magazines to read during those periods. Practice was continuous except for changing hands and for rest intervals.

The *S*s used were 73 males and 27 females with ages ranging from 18 to 25 yr., from the elementary psychology class at the University of Wisconsin. All *S*s were right handed and each was randomly assigned to one of the five conditions. The conditions were approximately equated for sex.

TABLE 1

DESIGN OF EXPERIMENT

Events	Conditions				
	A	B	C	D	E
Trials, right hand (prerest)	8	8	7	8	7
Rest in minutes	—	4	4	12	12
Warm-up trials on right hand	—	—	1	—	1
Trials, left hand (postrest)	8	8	8	8	8

RESULTS

Analyses of covariance were performed in order to compare postrest performance between conditions, with the scores corrected for ability on the right (pretransfer) hand. The first analysis was based on the mean of Trials 1 to 7 on the right hand and Trial 1 on the left (transfer) hand. With 4 and 94 *df*, *F* was 2.69 and significant at the .05 level. Table 2 contains both obtained and adjusted Trial 1 means. Tukey's (15) technique for comparing individual means was performed on the adjusted means and showed the difference to be between Cond. A (no rest) as a straggler, and the remaining four as a homogeneous group. The important thing to note is that although bilateral reminiscence is clearly indicated, the rest and warm-up conditions do not differ from one another on Trial 1.[5] That the low mean for Cond. A is not due to difficulty in changing hands is shown by the performance in Cond. C and E, the warm-up groups. The possibility was also considered that the warm-up effect may have carried over to Trial

TABLE 2

OBTAINED AND ADJUSTED SCORES FOR TRIALS 1 AND 2 ON THE LEFT HAND AND RESULTS OF TUKEY'S TEST

Cond.	Trial 1		Trial 2	
	Obt.	Adj.	Obt.	Adj.
A	.45	.47*	.75	.78*
B	.90	.98	1.22	1.37
C	1.04	.98	1.52	1.42
D	.91	.94	1.47	1.52
E	.86	.79	1.24	1.12

* Separated as stragglers from remaining homogeneous groups.

[5] The lack of a significant difference between the 4- and 12-min. conditions is not inconsistent with the unilateral rotary pursuit findings of Ammons (3) for 5 and 10 min. of interpolated rest and 3 min. of continuous prerest practice, or of Grice and Reynolds (6) for 5 and 10 min. of interpolated rest and 7.5 min. of distributed prerest practice. Whether or not a significant difference was obtained by Grice and Reynolds for the same time intervals in bilateral transfer is not immediately obvious from the data presented.

2. A similar analysis of covariance was performed for Trial 2, and with 94 and 4 *df*, *F* was 2.70 and significant at the .05 level. Again, Tukey's test on the adjusted means, as indicated in Table 2, shows the difference to be between Cond. A (no rest) as a straggler and the remaining four as a homogeneous group.

In order to test for unilateral warm-up, (i.e., significant improvement from left-hand Trial 1 to Trial 2, as compared to subsequent improvement) an analysis of variance was performed on the transfer-hand data comparing the successive differences between the first four trials. Only trial-difference scores were significant ($F = 4.99$; $df = 2, 190$; $P = .01$), which indicates that all five conditions were acting relatively the same over trials. The trial-difference means were .405, .151, and .134 for Trials 1–2, 2–3, and 3–4, respectively, with the mean for the difference between Trial 1 and Trial 2 accounting for the significant effect. Thus, unilateral warm-up was obtained even in the warm-up conditions, C and E, where bilateral warm-up was given an opportunity to replace it. In fact, left-hand warm-up also occurred in Cond. A in which no rest was given. These data support the finding that unilateral warm-up is relatively complete following 30 sec. of post-rest practice.

The data also provide a measure of unilateral reminiscence as well as of bilateral reminiscence. A simple analysis of variance was performed for all conditions, comparing the mean differences in performance between Trials 7 and 8 on the prerest or right hand. With 4 and 95 *df*, *F* was 12.13 and significant at the .001 level. The means of the differences for Cond. A through E were − .20, .23, 1.97, .28, and 2.00, respectively, and Tukey's test showed that Cond. C and E, the warm-up conditions, were separated from the remaining homogeneous group by a significant gap. Comparing these data to the results in Table 2 further substantiates previous findings of a relatively greater magnitude for unilateral reminiscence as compared to bilateral reminiscence.

Discussion

There are two major points of interest to be noted from the foregoing results: (*a*) Tests were made for bilateral warm-up and for bilateral reminiscence but only the latter was found. (*b*) Marked unilateral warm-up was found which was similar for all conditions, whether or not it could have been replaced by bilateral warm-up. In other words, practice on the right hand did not serve as warm-up for left-hand performance.

Thus, if warm-up is interpreted as establishment of set, the set involved here seems to be remarkably specific to performance on a given hand, much more so than the reactive inhibition which appears to underlie both unilateral and bilateral reminiscence.

Transfer of warm-up from one task to another depends on similarity of the tasks (**11**, p. 453). The absence of warm-up transfer in the present data suggests a lack of similarity between the left-hand and right-hand pursuit rotor tasks.

Transfer of work inhibition or reactive inhibition also appears to depend on task similarity (rather than effector localization) as shown in an inverted alphabet-printing task (**1**). The presence of moderate bilateral reminiscence (hence reactive inhibition) in the present data, therefore, suggests that the tasks do have some similarity.

These conclusions are not necessarily contradictory. Task similarity in motor learning involves more than the response product. It may be analyzable into component stimulus and response similarities. The results of this experiment indicate that the type of similarity required for transfer of set is different from that required for transfer of reactive inhibition. This experiment by itself does not identify these different types of similarities, but further experimental analysis of the role of effector involve-

ment, movements, and stimulus situations as affecting warm-up, work decrement, and reminiscence may do so. This approach may lead to better understanding of the nature and locus of the processes underlying motor activity.

For example, it might be pointed out that although bilaterally symmetrical effectors were used in this experiment, the movements of the two hands were not bilaterally symmetrical since they were not mirror images of each other. A pursuit rotor study might capitalize on the bilateral symmetry of the human anatomy by having the turntable rotate clockwise for right-hand practice and counterclockwise for left-hand practice. Using a different task (rapid tracing of a slot pattern with a stylus), Milisen and Van Riper (**12**) found that greater bilateral transfer of training occurred if the two hands traced the pattern in opposite rotary directions. Since their results could easily be interpreted in terms of task similarity, it appears probable that the task similarities involved in warm-up and work decrement could indeed be studied as proposed. The present study initiates such analysis by indicating that clockwise pursuit rotor performance is quite a different task for the two hands with regard to set, but moderately similar with regard to reactive inhibition.

Summary

This study was undertaken to investigate the effects of warm-up in bilateral transfer on a pursuit rotor task. There were five experimental conditions: a control condition with no rest, two conditions with 4 and 12 minutes rest, and two conditions with 4 and 12 minutes rest but with the final trial on the pretransfer hand interpolated as a warm-up trial immediately before practice on the transfer hand. Each group practiced a total of eight 30-second trials on each hand. Practice was continuous except for changing hands and for rest intervals.

Bilateral warm-up and bilateral reminiscence were predicted but only the latter was found. Unilateral warm-up on the transfer hand was a striking effect even where bilateral warm-up could have replaced it. Since warm-up occurs only unilaterally, set appears to be relatively effector-specific. On the other hand, the occurrence of both unilateral and bilateral reminiscence indicates reactive inhibition to be more general in its effects. It is suggested that transfer of warm-up and transfer of work decrement depend on different aspects of task similarity, the experimental analysis of which is feasible and may lead to better understanding of the processes underlying motor activity.

References

1. Albright, L. E., Borresen, C. R., & Marx, M. H. Reactive inhibition as a function of same-hand and opposite-hand intertrial activity. *J. exp. Psychol.*, 1956, **51**, 353–357.
2. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
3. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
4. Ammons, R. B., & Ammons, C. H. Bilateral transfer of rotary pursuit skill. *Amer. Psychologist*, 1951, **6**, 294. (Abstract)
5. Denny, M. R., Frisbey, N., & Weaver, J., Jr. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48–54.
6. Grice, G. R., & Reynolds, B. Effect of varying amounts of rest on conventional and bilateral transfer "reminiscence." *J. exp. Psychol.*, 1952, **44**, 247–252.
7. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
8. Irion, A. L. Reminiscence in pursuit-rotor learning as a function of length of rest and amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
9. Irion, A. L., & Gustafson, L. M. "Reminiscence" in bilateral transfer. *J. exp. Psychol.*, 1952, **43**, 321–323.
10. Kimble, G. A. Transfer of work inhibition in motor learning. *J. exp. Psychol.*, 1952, **43**, 391–393.
11. McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.

12. Milisen, R., & Van Riper, C. Differential transfer of training in a rotary activity. *J. exp. Psychol.*, 1939, **24,** 640–646.
13. Rockway, M. R. Bilateral reminiscence in pursuit-rotor learning as a function of amount of first-hand practice and length of rest. *J. exp. Psychol.*, 1953, **46,** 337–344.
14. Silver, R. J. Effect of amount and distribution of warming-up activity on retention in motor learning. *J. exp. Psychol.*, 1952, **44,** 88–95.
15. Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5,** 99–114.

(Received June 8, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

CONFLICT AND INFORMATION-THEORY VARIABLES AS DETERMINANTS OF HUMAN PERCEPTUAL CURIOSITY [1]

D. E. BERLYNE [2]

University of Aberdeen

The much needed integration of the theory of perception with general behavior theory would seem to require some account of the motivational factors underlying perception. In laboratory experiments, *E*s usually induce artificial motives in human *S*s by imposing specific tasks on them. Spontaneous human perceptual activities outside the laboratory, on the other hand, have been ascribed by some writers to such motives as "effort after meaning" (**2**) or "will to perceive" (**23**), reminiscent of the exploratory drives that are commonly invoked to explain exploratory behavior in animals (**4, 14, 18**).

The application of the concepts of "drive" and "drive-reduction" to this new area calls for some circumspection and some justification, as there may be more parsimonious ways of describing the facts. For example, the investigatory-reflex theory (**19, 21**) may perhaps be adequate to account for orienting movements which afford a clearer or intenser perception of some stimulus object to which an organism is already exposed. The theory is, however, inapplicable when the response resulting in optimal perception is not innate but arbitrarily selected by *E*, and the stimulus object in question is absent at the time the response is made.

These conditions obtained in the experiments to be reported in this article. Tachistoscopic exposures of visual figures resulted from pressing a finger key, and *S*s were free to expose themselves to each figure as many times as they wished. After each exposure, *S*s found themselves in darkness. Fluctuations in the probability of making another response can therefore be attributed to some internal stimulus condition or drive. If this probability turns out to depend on the nature of the figure just exposed and the number of times it has already been seen, these variables can be regarded as drive-conditions, influencing the strength of the drive. If the probability of a further response decreases as a particular figure is repeatedly exposed, the exposures can be said to bring about the satiation or reduction of the drive.

By analogy with such drives as hunger or thirst, which are commonly measured by amount of consummatory activity (cf. **3**), the number of exposures of a particular figure before *S* gave up responding was taken as an indication of the intensity of the drive aroused by that figure. The ultimate cessation of responding can be attributed, as with eating and drinking, both to a reduction in the energizing effect of the drive and to changes in internal stimuli. Of these, the latter would seem to be more important. It would not seem to be a matter of extinction in either case, as the response is immediately restored to its original strength when motivation is renewed.

A drive which is reduced by perception has been called *perceptual*

[1] The author is indebted to Mr. S. E. Burgess, who carried out a pilot study with children and is responsible for the preparation of most of the stimulus material.

[2] Now temporarily at the Center for Advanced Study in the Behavioral Sciences.

curiosity and provisionally distinguished from *epistemic curiosity*, which is reduced by the acquisition of knowledge (**5**). One class of variable that appears to increase perceptual curiosity in lower animals is the complexity of a stimulus situation (**8, 9, 12, 22**). This is related to relative entropy, the inverse of redundancy (**20**), which Attneave (**1**) has shown to be an important attribute of visual figures. One of the principal determinants of epistemic curiosity, according to a theory proposed by the writer (**5, 6, 7**), is the degree of learned conflict between the symbolic responses aroused by a stimulus situation, a variable which is related to the information theorist's entropy or uncertainty (**20**). An important case of such conflict is *incongruity-conflict*, aroused by a stimulus pattern with characteristics which *S* has been trained to regard as incompatible. Another is *surprise-conflict*, produced when a stimulus pattern fails to confirm an expectation evoked by what preceded it. Secondly, the theory implied that epistemic curiosity will increase with the number of previous gratifications of the drive in similar situations. If all these factors can be shown to affect human perceptual curiosity, then the grounds for differentiating between perceptual and epistemic curiosity will become more tenuous.

Method

Apparatus

A tachistoscope (a simplified version of the one described in **15**), operating with two short-decay phosphor fluorescent tubes and taking 6-in. square cards, was connected with a finger key. Whenever the key, which resembled the bar of a Skinner box, was depressed, the card in the tachistoscope became visible for .14 sec.

Subjects

Sixteen undergraduates from the first- and second-year psychology classes served as *Ss*.

Procedure

The *S* sat in a darkened room, 4 ft. from the tachistoscope, with his right hand holding the key. The use of the key was explained and demonstrated with two sample cards reserved for this purpose. He was told that the experiments were intended solely to show how interesting certain pictures were. The *E* would insert a card in the tachistoscope, and *S* was to begin pressing the key when *E* said "Now." The *S* was to continue to press the key for as long as he wanted to continue seeing the card. When he no longer wanted to see it, he was to say "Yes," and *E* would insert a new card. It was emphasized that no questions about the cards would be asked at any time.

Design

Each *S* underwent four experiments, which followed one another without interruption, used the same procedure, and differed only in stimulus material. Different *Ss* had the four experiments in four different sequences so that the design took the form of four replications of a counterbalanced 4 × 4 latin square (**10**).

Stimulus Material

Experiment 1 (*Incongruity-conflict*).—There was a series of seven pictures of Animals and a series of seven pictures of Birds (Fig. 1). Cards 2 and 4 of the Animals series and 3 and 5 of the Birds series bore *incongruous* pictures. Half of the *Ss* received the Animals series first and the other half the Birds series first.

Experiment 2 (*Meaningful sequence*).—There was a series of six cards which began with a circle and developed by progressive addition of details into a picture of a Bear and a similar series of six cards changing from a circle into a picture of a Clown (Fig. 2). Each *S* received one series in this meaningful order and the other series in a random order. The four possible sequences—Bear meaningful Clown random, Clown meaningful Bear random, Bear random Clown meaningful, and Clown random Bear meaningful—were used with equal numbers of *Ss*.

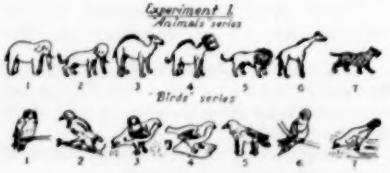

Fig. 1. The stimulus material, Exp. 1.

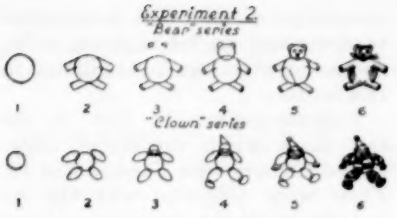

FIG. 2. The stimulus material, Exp. 2.

Experiment 3 (Surprise-conflict).—There were 12 cards bearing geometrical patterns made up of colored spots (Fig. 3). The patterns of Cards 1–6 were composed of red triangles, those of Cards 7–11 of green circles, and that of Card 12 of violet squares.

Experiment 4 (Relative entropy).—There were three series of cards each bearing patterns differing in redundancy. As Attneave (1) shows, redundancy varies directly with orderly

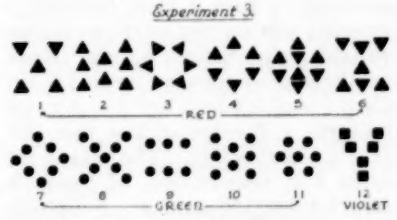

FIG. 3. The stimulus material, Exp. 3.

arrangement, with symmetry and with similarity between parts of a figure, and inversely with changes in contour. Series A contained a card bearing nine crosses in a regular matrix pattern and a card bearing nine crosses in an irregular arrangement (Fig. 4). Series B consisted of four cards bearing a circle, a square, an octagon, and an irregular closed curve, all of approximately equal perimeter. These figures have, respectively, zero, four, eight, and an infinite set of points at which the curvature changes and hence represent progressively

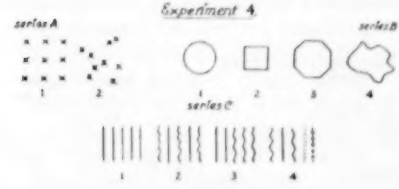

FIG. 4. The stimulus material, Exp. 4.

smaller degrees of redundancy. The four cards of Series C present patterns of five straight lines, of two straight and three wavy lines arranged symmetrically, of two straight and three wavy lines arranged asymmetrically and of five lines of different sorts. Redundancy therefore increases from the first card to the last.

Two sequences, based on random numbers, were worked out for all 10 cards, and each sequence was used with half the *S*s.

RESULTS AND DISCUSSION

Determinants of Curiosity

Previous gratifications of the drive.—The mean numbers of responses per card were worked out for *S*s by experiments and are displayed in Table 1. They were then subjected to an analysis of variance, with the results shown in Table 2. It will be seen that the effect of the temporal position of an experiment is significant. When this effect is analyzed further, it is found that the linear and quadratic components of the trend are significant but that the deviation from the quadratic trend is not. This means that the data do not depart significantly from the best-fitting quadratic curve and that the trend represented by this curve is significant. The curve happens to be a negatively accelerated rising one, similar in shape to the usual learning curve. The results are therefore compatible with the assumptions of the theory (**5, 6**), according to which curiosity will increase with the num-

TABLE 1

MEAN RESPONSES PER CARD BY EXPERIMENTS

Exp.	Mean	*SD*	Order	Mean	*SD*
1	3.8	1.32	I	2.4	.84
2	2.4	.91	II	2.6	1.20
3	2.1	.80	III	3.0	1.20
4	2.7	.92	IV	2.7	1.40
All	2.8	.84			

TABLE 2

BETWEEN-EXPERIMENTS ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Sequences	3	1.088	
Ss within sequences	12	2.806	
Total between *Ss*	15	2.462	10.42**
Experiments	3	9.106	38.58**
Temporal positions:			
a. linear component	1	1.435	6.08*
b. quadratic component	1	1.331	5.64*
c. deviation from quadratic trend	1	0.740	3.14
Temporal positions: total	3	1.169	4.95*
Residual (error)	42	0.236	
Total	63		

* Significant at .05 level.
** Significant at .001 level.

ber of previous opportunities to gratify it in similar circumstances.

Incongruity-conflict.—In Exp. 1, the mean numbers of responses per card were 5.8 for the four incongruous pictures and 3.0 for the 10 nonincongruous pictures. As the analysis of variance summarized in Table 3 shows, the difference is highly significant, and so the predicted importance of incongruity-conflict is vindicated.

Meaningfulness of sequence.—Although meaningfulness of sequence was not one of the variables figuring in the theory, it was tested in Exp. 2 as a factor likely to be influential. The means were, however, 2.4 responses per card for both meaningful-sequence and random-sequence series, and so no effect can be attributed to this factor.

Surprise-conflict.—In Exp. 3, the two cards which introduced unexpected changes were Cards 7 and 12. These were compared with the remainder, with the exception of Card 1. Card 1 was disregarded because of its equivocal status at the beginning of the series. The two surprise cards produced a mean of 2.4 responses and the nine nonsurprise cards a mean of 2.0 responses. As the analysis of variance in Table 4 shows, the difference, which follows the predicted direction, is significant.

Relative entropy.—The three series of cards in Exp. 4 each provided a separate test for the relative-entropy variable. In Series A, the mean numbers of responses were 2.6 for Card 1 and 4.2 for Card 2. A t test leads to a P-value of less than .05 ($t = 2.43$, $df = 15$). The means for cards 1–4, respectively, of Series B were 1.2, 1.6, 1.8, and 3.2. For Cards 1–4, respectively, of Series C, they were 1.7, 2.6, 2.8, and 3.3. In both of these series, therefore, the means followed exactly the predicted order (out of 24 possible permutations in each case). Ranking analyses of variance demonstrate that the differences, for both series, reach the .01 level of significance ($\chi^2_r = 18.47$ for

TABLE 3

ANALYSIS OF VARIANCE: EXP. 1

Source	*df*	*MS*	*F*
Incongruous/Non-incongruous I/NI	1	382.00	(*a*) 32.73**
Ss	15	26.20	(*b*) 6.44**
Ss × I/NI (*a*)	15	11.67	(*b*) 2.87*
Residual (*b*)	192	4.07	
Total	223		

* Significant at .01 level.
** Significant at .001 level.

TABLE 4

ANALYSIS OF VARIANCE: EXP. 3

Source	*df*	*MS*	*F*
Surprise/Nonsurprise (S/NS)	1	4.00	6.35*
Ss	15	6.80	10.79**
Ss × S/NS	15	.47	
Residual (error)	144	.63	
Total	175		

* Significant at .01 level by one-tailed test.
** Significant at .001 level.

TABLE 5

ANALYSIS OF VARIANCE: EXP. 2

Source	*df*	*MS*	*F*
Amounts of detail (D)	2	19.50	(*a*) 9.90**
Ss	15	10.13	(*b*) 13.51**
Ss × D (*a*)	30	1.97	(*b*) 2.63**
Residual (*b*)	144	.75	
Total	191		

** Significant at .001 level.

Series B and 15.69 for Series C, $df = 3$).

Absolute entropy.—The variable tested in Exp. 4 was relative entropy or uncertainty. This is the ratio of the amount of information received from a figure to the maximum that could be received from a figure with the same quantity of material. In other words, it is a measure of the extent to which parts of the figure are unexpected in view of the nature of the other parts or, more briefly, a measure of the amount of variety in the figure.

All three series consisted of figures with approximately equal amounts of material. Experiment 4, therefore, showed that curiosity increases with relative entropy if maximum entropy is held constant. The figures in each series thus varied both in relative and in absolute entropy, and the question arises whether the *absolute* amount of information as such has an influence on curiosity.

Affirmative evidence on this question is provided by Exp. 2. It will be seen that the cards of both series numbered 1, 2–5 and 6 form three sets differing markedly in amount of nonrepetitive detail. The mean responses per card for these three sets are 1.6, 2.3, and 3.1, respectively, and, as Table 4 shows, the F value for this comparison is significant. The differences between figures varying in amount of detail can be described in terms of statistical information theory (**20**), if the parts of the figures are regarded as signals. But the concept of "structural information" (**17**) is clearly also relevant.

Alternative Theories

The Pavlovian concept of an "investigatory reflex" (**19, 21**) cannot account for our results, as the response was not a receptor adjustment but an arbitrary learned movement and, furthermore, the visual stimulus investigated was not present until the response had been performed. Estes (**13,** p. 11), referring to experiments which purport to demonstrate learning under the influence of an exploratory drive, claims that Guthrie's contiguity theory could equally well handle the facts. It is not easy to see how this theory could handle our data, without assuming some internal stimulus condition which would amount to much the same as a curiosity drive.

Comparison with Children

Burgess (**11**) performed Exp. 1 and slightly modified versions of Exp. 2 and 3 with 18 5-yr.-old children and with a .014-sec. exposure. The principal differences from our results were: (*a*) the children responded many more times per card (mean = 12.1, $SD = 7.75$); and (*b*) no effect of incongruity or surprise was manifested. In order to ascertain whether the first difference was due to the age of the children or to the shorter exposure, four *Ss* from the fourth-year psychology class were put through Exp. 1–4 with a latin-square design and a .014-sec. exposure. Their mean number of responses per card for Exp. 1–3 came to 2.3, with an SD of .70. A Cochran-Cox test (**16,** pp. 97–98) shows the difference between their mean and that of the children to be significant at the .01 level. One can therefore conclude that the children's greater resistance to satiation was not due to the shorter exposure alone. Without further experimentation, it is not possible to reach any conclusion with respect to the other difference. The fact that the children

exhibited more curiosity than the adults is in harmony with Welker's finding (22) that young chimpanzees indulge in more exploratory manipulation of objects than their elders.

Summary

Sixteen undergraduate *Ss* each took part in four experiments in which a key-pressing response produced tachistoscopic exposures of visual figures. The *Ss* were free to have as many exposures of each figure as they wished, and the number of responses performed was regarded as an indication of the intensity of the curiosity aroused by a particular figure. A counterbalanced latin-square design was used.

The purpose of the experiments was to test some variables which might be expected to influence curiosity, most of them being suggested by studies of exploratory behavior in lower animals or by the writer's theory of human epistemic curiosity. The following properties of figures were found to increase curiosity significantly: (*a*) incongruity, (*b*) surprisingness, (*c*) relative entropy, and (*d*) absolute entropy. The mean numbers of responses per card in successive experiments were significantly different, the data being adequately accounted for by a rising, negatively accelerated quadratic curve.

Theoretical questions raised by the concept of a curiosity drive, alternative ways of accounting for the phenomena, and data obtained from using the same experimental technique with children are discussed.

References

1. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61,** 183–193.
2. Bartlett, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.
3. Berkun, M. B., Kessen, M. L., & Miller, N. E. Hunger-reducing effects of food by stomach fistula versus food by mouth measured by a consummatory response. *J. comp. physiol. Psychol.*, 1952, **45,** 550–564.
4. Berlyne, D. E. Novelty and curiosity as determinants of exploratory behaviour. *Brit. J. Psychol.*, 1950, **41,** 68–80.
5. Berlyne, D. E. A theory of human curiosity. *Brit. J. Psychol.*, 1954, **45,** 180–191.
6. Berlyne, D. E. An experimental study of human curiosity. *Brit. J. Psychol.*, 1954, **45,** 256–265.
7. Berlyne, D. E. Knowledge and stimulus-response psychology. *Psychol. Rev.*, 1954, **61,** 245–254.
8. Berlyne, D. E. The arousal and satiation of perceptual curiosity in the rat. *J. comp. physiol. Psychol.*, 1955, **48,** 238–246.
9. Berlyne, D. E., & Slater, J. Perceptual curiosity, exploratory behavior, and maze learning. *J. comp. physiol. Psychol.*, 1957, **50,** 230–234.
10. Bugelski, B. R. A note on Grant's discussion of the Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1949, **46,** 49–50.
11. Burgess, S. E. A study of curiosity as motivation for a key-pressing response in children. Unpubl. Ed.B. thesis, Univer. of Aberdeen, 1956.
12. Butler, R. A. Incentive conditions with influence exploration. *J. exp. Psychol.*, 1954, **48,** 19–23.
13. Estes, W. K. Learning. *Annu. Rev. Psychol.*, 1956, **7,** 1–38.
14. Harlow, H. F., Harlow, M. K., & Meyer, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.*, 1950, **40,** 228–234.
15. Humphrey, G., Dawe, P. G. M., & Mandell, B. New high-speed electronic tachistoscope. *Nature*, 1955, **176,** 1–9.
16. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Cambridge, Mass.: Houghton Mifflin, 1953.
17. Mackay, D. M. Quantal aspects of scientific information. *Phil. Mag.*, 1950, **41,** 289–311.
18. Montgomery, K. C. The role of the exploratory drive in learning. *J. comp. physiol. Psychol.*, 1954, **47,** 60–64.
19. Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
20. Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. of Ill. Press, 1949.
21. Sokolov, E. N. The higher nervous activity and the problems of perception. *Proc. XIV Int. Cong. Psychol.*, 1955, 134–135.
22. Welker, W. I. Some determinants of play and exploration in chimpanzees. *J. comp. physiol. Psychol.*, 1956, **49,** 84–89.
23. Woodworth, R. S. Reinforcement of perception. *Amer. J. Psychol.*, 1947, **60,** 119–124.

(Received May 28, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

THE VERTICAL–HORIZONTAL ILLUSION AND THE VISUAL FIELD

THEODOR M. KÜNNAPAS

University of Stockholm, Sweden

The overestimation of the vertical direction as compared with the horizontal direction of equal length has been known for more than a hundred years. Since 1851, when Fick (**1, 2**) first drew attention to the vertical illusion, different attempts have been made to explain the phenomenon, e.g., the eye-movement theory of Wundt (**9, 10**), the perspective theory of Thiéry (**8**), the empathy theory of Lipps (**6, 7**) and the anisotropic space theory of Koffka (**3**). None of these theories has succeeded in giving an explanation that is quite satisfactory.

In the present paper another attempt is made to explain the vertical-horizontal illusion. In a previous report (**5**), it was found that the size of the frame influences the apparent length of a line enclosed within this frame. It was pointed out that an increase in the size of the frame reduces the phenomenal length of the line drawn on it. It may be assumed that the visual field is one of the visual frames which may influence the way in which things are seen. The visual field has really the form of a horizontal ellipse. If, in the center of this field, there are two lines of equal length, one vertical and the other horizontal, the distance *a* (see Fig. 1) from the vertical line to the boundary of the visual field is shorter than the distance *b* from the horizontal line to the boundary. Thus, according to the previous report (**5**), the vertical direction may be expected to appear longer than the horizontal.

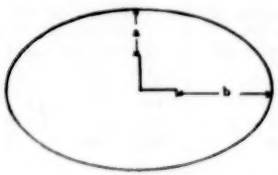

FIG. 1. Schematic visual field.

The general hypothesis is that the vertical direction is overestimated because the visual field has the form of an ellipse which is extended in the horizontal direction. On the basis of this hypothesis, several testable predictions can be made; and one of these will be tested in the present paper. If the overestimation of the vertical direction is due to the elliptical form of the visual field with its greater horizontal axis, it must disappear in complete darkness where the visual field has no distinct boundary. But owing to force of habit in connection with the visual field, one cannot expect that its influence will suddenly disappear. Therefore, if the overestimation of the vertical in the dark situation is significantly reduced, this will support the hypothesis.

METHOD

Stimulus pattern.—According to a previous paper (**4**), the inverted T figure contains not only the vertical illusion but also an illusion due to the overestimation of the dividing line. For this reason an L-shaped stimulus pattern was used.

Apparatus.—The apparatus consists of a circular white surface with a radius of 300 mm. (see Fig. 2). On the central part of this surface the stimulus pattern can be projected from the rear. It consists of two luminous lines (black in Fig. 2), seen through two slits .5 mm. wide, which are themselves invisible on the round white surface. One of these lines is horizontal

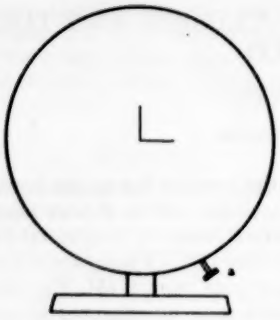

FIG. 2. The apparatus.

and the other vertical. The horizontal line is the standard stimulus S_s. It has a constant length of 50 mm. The vertical line is the variable stimulus S_v. Its length may be varied from 35 to 60 mm. by turning a little knob (*a*). Several turns are required to get a noticeable adjustment. The size of S_v can be read from a millimeter scale behind the screen. The stimulus pattern is exposed by switching on the light.

Procedure.—The *O* sat in front of the apparatus and rested his chin on a support, in a constant position. The distance between *O*'s eyes and the figure in the center of the round surface was 400 mm., corresponding to a visual angle of 7°. The figure itself, which *O* had to fixate, was at eye level.

The method of average error was employed. The *O* saw the figure only when he was adjusting S_v. After the adjustment, the figure was switched off by *E*. When *E* had brought S_v to a new position which was obviously shorter or longer than S_s, the figure was exposed again. During the adjustment of S_v, *O* tried to fixate the center of the pattern where the vertical and the horizontal lines intersect.

The stimulus pattern was viewed with both eyes. In preliminary trials *O* was made familiar with the experimental situation.

Experimental situations.—In the Light situation, the room in which the experiments were conducted was illuminated by a 150-w. lamp suspended from the ceiling behind *O*. In the Dark situation, this room was completely dark during the adjustment of S_v. After each adjustment, the room was illuminated again in order to prevent dark adaptation. The sequence of the Light and Dark situations was LDDL or DLLD.

Observers.—The *O*s were 20 university students in psychology, 10 men and 10 women. All *O*s had normal visual acuity; in four cases sight was corrected by eyeglasses. Ten *O*s began with the Light condition and 10 with the Dark. Each *O* made 32 adjustments, 16 ascending and 16 descending trials, in each experimental situation. Thus, each *O* made 64 adjustments and the group made a set of 1280 adjustments.

TABLE 1

LENGTH OF THE VERTICAL LINE (IN MILLIMETERS) WHICH APPEARS EQUAL TO HORIZONTAL LINE OF 50 MM.

Cond.	Ascending Mean	Descending Mean	Total	
			Mean	*SD*
Light	45.93	46.99	46.46	1.92
Dark	47.02	48.20	47.61	2.18

RESULTS

The effects of conditions and types of series are summarized in Table 1. It is clear that, in the Dark situation, the overestimation of the vertical line decreased, the magnitude of the illusion being 4.8% instead of 7.1%. The direction of this difference was significant at the .01 level or better in 16 of the 20 *O*s; was present in 2 additional *O*s, but not significantly; and was reversed in 2 *O*s, but not significantly.

The results of an analysis of variance of these data are given in Table 2. Because of the significant interactions (conditions × *O*s and types

TABLE 2

ANALYSIS OF VARIANCE OF THE DATA

Source	*df*	*MS*	*F*	*P*
Conditions (A)	1	408.75	417.09	<.001
Series (B)	1	374.90	382.55	<.001
Observers (*O*)	19	198.52	202.57	<.001
A × B	1	1.51	1.54	
A × *O*	19	7.64	7.80	<.001
B × *O*	19	8.30	8.47	<.001
A × B × *O*	19	1.17	1.19	
Within	1200	.98		
Total	1279			

of series × *Os*), these interactions were used for testing the significance of the main effects, as appropriate. The reduction of the overestimation of the vertical direction in the dark condition was significant above the .001 level. It will be noted that, while the effect of Series (ascending or descending) is significant, the interaction of the Light and Dark situation and the type of series is not significant. Thus, the hypothesis is supported by the experiment.

Summary

A new attempt is made to explain the overestimation of the vertical direction, on the basis of the hypothesis that the vertical direction, as compared with the horizontal direction, is overestimated because the shape of our visual field is roughly elliptical or oval which is extended in the horizontal direction. This hypothesis leads to the prediction that the vertical illusion must decrease significantly in complete darkness, where the visual field has no distinct boundaries.

Experiments in Light and in Dark conditions were conducted with the same *Os*. The method of average error was employed with an L-shaped stimulus pattern. It was found that in the Dark condition, the overestimation of the vertical direction was reduced by 30%. The reduction was significant above the .001 level. Thus, the hypothesis is supported by the experiment.

References

1. Fick, A. *De errore quodam optico asymmetria bulbi effecto.* Marburg: J. A. Kochii, 1851.
2. Fick, A. Erörterung eines physiologisch-optischen Phänomens. *Z. ration, Med.*, 1852, **2**, 83–94.
3. Koffka, K. *Principles of Gestalt psychology.* London: Routledge & Kegan Paul, 1950.
4. Künnapas, T. M. An analysis of the "vertical-horizontal illusion." *J. exp. Psychol.*, 1955, **49**, 134–140.
5. Künnapas, T. M. Influence of frame size on apparent length of a line. *J. exp. Psychol.*, 1955, **50**, 168–170.
6. Lipps, T. Aesthetische Faktoren der Raumanschauung. In *Beiträge zur Psychol. und Physiol. der Sinnesorgane*, published as Festgruss to Helmholtz. Hamburg & Leipzig: Leopold Voss, 1891. Pp. 217–307.
7. Lipps, T. Raumaesthetik und geometrisch-optische Täuschungen. *Schriften der Gesellschaft f. psychol. Forsch.*, No. 9 & 10. Leipzig: J. A. Barth, 1897.
8. Thiéry, A. Ueber geometrisch-optische Täuschungen. *Philos. Stud.*, 1896, **12**, 67–125.
9. Wundt, W. *Beiträge zur Theorie der Sinneswahrnehmung.* Leipzig & Heidelberg: C. F. Winter, 1862.
10. Wundt, W. Die geometrisch-optischen Täuschungen. *Abh. sächs. Ges. Wiss., math.-phys. Cl.*, 1898, **24**, 53–178.

(Received June 11, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

DIFFERENTIAL EYELID CONDITIONING AS A FUNCTION OF STIMULUS SIMILARITY AND STRENGTH OF RESPONSE TO THE CS [1]

MALCOLM D. GYNTHER [2]

Duke University

In the various theories (**7, 14**) of discrimination learning derived from Pavlov's (**13**) early explanation, two general sorts of variable are of great importance. They are the psychophysical difference between the stimuli to be discriminated and the absolute strengths of excitatory and inhibitory tendencies conditioned to the positive (reinforced) and negative (nonreinforced) stimuli. The experiment reported in this investigation manipulated these variables in a number of ways in order to test several predictions which follow from such theorizing. Basic to these theories are the following five assumptions: (*a*) as a function of reinforcement, there develops an increasing excitatory potential associated with a positive stimulus (CS +); (*b*) with nonreinforcement, an increasing inhibitory potential associated with a negative stimulus (CS −) develops; (*c*) the total amount of excitation is greater than the total amount of inhibition; (*d*) both the excitatory and inhibitory tendencies generalize to adjacent stimulus values; and (*e*) response strength at any point on the stimulus continuum is an algebraic function of the magnitude of the excitatory and inhibitory tendencies. These assumptions, with certain supporting hypotheses, are capable of mediating the prediction of a substantial number of facts about discrimination. Some of these have been presented in the writings of Spence (**15, 16, 17**), Grice (**4**), Ehrenfreund (**2**), Margolius (**12**), Antoinetti (**1**), and Kendler (**9**). Several other predictions to which the experimental section of this paper is directed are described below.

1. In a discrimination situation where CS + and CS − are similar but distinguishable, performance to CS + as a function of the number of reinforced trials should be lower than performance in an ordinary conditioning situation where there are no negative trials. The reason for this is that extinctive inhibition associated with CS − should generalize to the positive stimulus in the discrimination situation and reduce performance to it. Such extinctive inhibition, of course, does not exist in the ordinary conditioning situation. The literature seems to contain no direct tests of this prediction. Data are available from different experiments which investigated the eyelid response (**19, 20**), but even gross comparisons are difficult to make due to the multitude of experimental conditions and response measures used.

2. A second prediction from this type of discrimination learning theory is the obvious one that increasing the similarity of positive and negative stimuli will increase the difficulty of establishing differential responses to the two stimuli. Two stimuli which are different psychophysically, in Hull's conception, are widely separated in terms of the generalization gradient as well. In such a

[1] This article is based upon a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Duke University. The author wishes to express his appreciation to Dr. Gregory A. Kimble whose continuous support and guidance contributed much to the formulation and execution of this research.

[2] Now at South Carolina State Hospital, Columbia, S. C.

case there will be little generalization of excitation to the negative stimulus and little generalization of inhibition to the positive one. This defines a situation where discrimination learning should readily occur. Bringing the stimuli closer together increases the amount of such generalization and, theoretically, should influence behavior in two ways: (*a*) the level of performance to CS + should decrease for the reasons detailed above, and (*b*) there should be a corresponding decrease in the differential response to the two stimuli. The logical extreme of bringing the two stimuli closer and closer together is the partial-reinforcement situation which of course produces no discrimination. Theoretically it should also lead to a considerable reduction in over-all level of responding. The pioneer experiment (**8**) suggested that performance in a partial-reinforcement situation is almost equal to that in the continuous reinforcement situation; however, more recent experiments (**3, 11**) have indicated that performance is a positive function of the percentage of reinforced trials. Again, however, performance under conditions of discrimination learning, partial reinforcement, and continuous reinforcement has never been directly compared. The present study provides data on this point.

3. Another cluster of variables theoretically influencing discrimination learning involves the relative strength of conditioning to CS + and CS −. Exactly what predictions one makes for this variable depends upon the manner in which these relative strengths are achieved. Consider first the case where *E* increases the relative strength of conditioning to CS + by giving relatively more trials with it than with CS −. According to Hull's theory, the strength of conditioning to CS + depends only upon the number of reinforced trials and is specifically independent of the number of nonreinforced trials with CS −. *Performance*, on the other hand, will depend considerably upon the number of such nonreinforced trials. If the number is large, the inhibition developed will accumulate and depress performance both to CS − which generated it and to CS + through generalization. Decreasing the number of negative trials relative to the number of positive ones will decrease inhibition for two reasons: (*a*) fewer responses generate less inhibition, and (*b*) these trials will be separated more in time on the average, permitting greater dissipation of inhibition. These considerations lead to the prediction that the level of responding to CS + will increase as one decreases the relative number of applications of CS −. Just what happens to *discrimination* under these circumstances depends upon the precise shape of the generalization gradient for different strengths of inhibition. If one makes the simplest assumption, that the generalization gradient for weak inhibition is exactly parallel to that for a greater amount of inhibition, the increase in the proportion of positive trials should have no effect at all upon the difference between the number of responses to CS + and CS −. This prediction follows from the fact that, on the assumption of parallel gradients, the increased amount of inhibition by comparison with a smaller amount of inhibition subtracts exactly as much more from performance to CS + as it does from CS −. Thus, at the same time that performance to CS + increases, there should be no change in differential conditioning, since there will be just as much increase to CS −.

4. Another way to increase the strength of conditioning to CS + in comparison with CS − is to condition *S*s to CS + prior to the introduction of the discrimination-learning procedure. In such preconditioning, little or no generalized extinctive inhibition will be developed. The introduction of trials with CS −, however, does lead to the development of inhibition which will generalize and reduce performance to CS +; generalization from CS + to CS −, at the same time, will lead to the occurrence of CR's to the latter stimulus. Responses to CS − will also be favored by context cues which are common to CS + and CS −. As discrimination training proceeds, responses to CS −

will tend, at first, to increase with responses to CS +; then as responses to the context cues extinguish, there will be a decrease in the number of CR's to CS −. The mechanisms involved here have been worked out in detail by Hull (7). Thus, after preconditioning, two effects are anticipated: (*a*) an initial reduction in performance to CS +, and (*b*) a period during which responses to CS + and CS − increase together before discrimination becomes evident. Antoinetti's experiment, described by Hull (7), supports both of these predictions. That of Spence and Beecroft (18), although they do not mention the fact, supports the first prediction, but neither supports nor questions the second.

Method

Apparatus.—The apparatus was that described by Kimble and Dufort (10). The *S* sat at a table with his chin and forehead supported by an adjustable headrest, looking into a booth which kept *E* and the timing apparatus out of his view and provided fairly constant conditions of general illumination. At the back of the booth, 16 in. in front of *S*, was a box with two ¼-in. holes set 2 and ⅞ in. apart. The CS + and CS − consisted of the illumination of one of these light apertures. Each CS was a dim (approximately 1.7 ml.) white light. The CS + was always the right-hand light. For the conditioning and partial-reinforcement procedures, the left-hand aperture was covered with a strip of black masking tape. For one discrimination procedure involving a greater physical difference between stimuli, the left-hand aperture (CS −) was covered by a strip of red plastic which had a peak transmission of about 635 mu.

Both apertures were open and of the same hue (white) under all other experimental conditions. The UCS was an air puff delivered to the right eye through a $\frac{3}{64}$-in. nozzle. It was magnetically released and provided by the fall of a 90-mm. column of mercury in a manometer. The duration of the CS (250 msec.), interstimulus interval (500 msec.), and the intertrial interval were controlled by a constantly running motor which activated a series of switches.

A very light false eyelash made of aluminum foil attached to *S*'s right eyelid interrupted the focal point of a light beam directed at the element of a photocell. The beam was focused on the eyelash about 1¼ in. from the eye. When *S* blinked approximately .075 in., the light reached the photocell and activated a circuit controlling the recording mechanism. The occurrence of a blink within the CS-UCS interval activated a light to indicate a CR.

Subjects.—The *S*s were 100 undergraduate students (50 men and 50 women) at Duke University, all of whom participated in this experiment as part of the laboratory requirement in the general psychology course. An additional 23 *S*s were eliminated for various reasons, such as extremely rapid blink rate, lack of cooperation, responding more than four times during adaptation trials and failing to meet other criteria.

Procedure.—Instructions were the same as those used by Kimble, Mann, and Dufort (11), modified where necessary for the discrimination-learning procedures. After the instructions had been presented and *S* had been properly adjusted to the apparatus, the CS + was presented alone 10 times. The *S*s who gave more than four responses to the CS on these trials were discarded. Following a 1-min. rest, *S*s received training trials with an intertrial interval of 15 sec., with the exception of the conditioning group whose trials were separated by 30 sec.

Design.—In this experiment, there were six groups of *S*s, consisting of a basic control group containing 30 *S*s (15 men and 15 women) run throughout the experiment, and five experimental groups containing 14 *S*s each (7 men and 7 women) run successively. It seemed essential to run the control *S*s in this fashion in order to eliminate possible systematic errors related to the time of running the experiment. The Control Group received 100 training trials with 1-min. rests after Trials 35 and 70. For this group, the CS − was haphazardly alternated with the CS +, the ratio of positive to negative trials being 1:1. Randomization was restricted in that: (*a*) no more than two positive or two negative trials were allowed to occur in a row, and (*b*) within each block of 10 trials, 5 trials were positive and 5 were negative.

Experimental Group 1 received 50 conditioning trials to CS + with a 30-sec. intertrial interval in order to keep the spacing of trials with CS + the same for both it and the Control Group. Comparison of Group 1 and the Control Group is to test the hypothesis that *S*s trained by classical conditioning procedures will perform significantly better to CS + than *S*s conditioned by a discrimination-learning procedure.

Experimental Group 2 received 67 training trials with a 1-min. rest after Trial 35. For this group, the ratio of CS + trials to CS − trials was about 3:1. Of the 67 trials, 50 were

reinforced and 17 were nonreinforced in random sequence, except that no more than five CS + nor more than two CS − were allowed to occur successively. The reason for this particular number of CS + and CS − is to hold the number of presentations of CS + the same as for the Control Group. The purpose of this experimental group is to test the predictions that: (*a*) *Ss* trained by the 3:1 procedure will give more CR's to both CS + and CS − than the Control Group, and (*b*) *Ss* given 3:1 training will discriminate no better than the Control Group.

Experimental Group 3 was a preconditioned group which received 20 conditioning trials, followed by 100 discrimination trials for those *Ss* who gave three or more CR's during preconditioning. This criterion was chosen as Kimble and Dufort (**10**) have shown that an *S* who has given three CR's in conditioning has reached the asymptote of the learning curve. Seven *Ss* were discarded for failure to meet this criterion in 20 trials. The 100 discrimination trials followed exactly the same pattern of rests, reinforcements, and nonreinforcements as with the control group. The purpose of the preconditioned experimental group is to test the hypotheses that: (*a*) *Ss* will give fewer responses to CS + on the first several reinforced trials in discrimination training as compared to their previous level of response to CS + during the last 10 preconditioning trials, (*b*) *Ss* trained by this procedure will give nearly as many responses to CS − as to CS + early in training (Trials 1 to 20), and (*c*) preconditioned *Ss* will begin to show more CR's to CS + than to CS − later in training (Trials 31 to 50).

Experimental Group 4 received 100 training trials with the same pattern of reinforcements, etc., as the Control Group. For this group, a strip of red plastic was placed over the CS − to increase the difference between stimuli. The purpose of this group is to test the predictions that *Ss* trained under these conditions will: (*a*) respond at a higher level to CS + than the Control Group, and (*b*) discriminate better than the Control Group.

Experimental Group 5 received 100 training trials, all of which were on the right-hand CS which was haphazardly positive and negative in the same pattern as the Control Group. This partial-reinforcement procedure represents the logical extreme in increasing the similarity between stimuli. The purpose of this experimental group is to test the hypotheses that under these conditions: (*a*) discrimination will not occur, and (*b*) performance to CS.+ will be lower than in the case of the Control Group.

Results [3]

The major results of the experiment are presented in a series of five graphs which present learning curves for each experimental group and the Control Group. In each graph the baseline is in units of 10 *reinforced* trials, and the ordinate is mean percentage of conditioned responses.

The data most relevant to the hypothesis which Group 1 was designed to test are presented in Fig. 1, which compares performance in a basic discrimination situation with that in conditioning. Inspection of the curves suggests that the main experimental prediction was supported. That is, the performance of the conditioning group, in terms of CR's to CS +, is at every point better than that of the discrimination (control) group. As would be predicted from the similarity of CS + to CS −, discrimination occurs but is relatively poor. These interpretations are supported by statistical analysis. A t test for significance of difference between group means on the last block of trials ($t = 2.69$, $P = .01$) indicates that the introduction of negative trials results in a decrement in performance to CS +. Within the discrimination group, the difference between responses to CS + and CS − on the last block of trials showed that discrimination does occur ($t = 4.18$, $P = .001$). However, the difference between number of re-

[3] Data for individual *Ss* and analysis of variance tables have been deposited with the American Documentation Institute. Order Document No. 5166 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

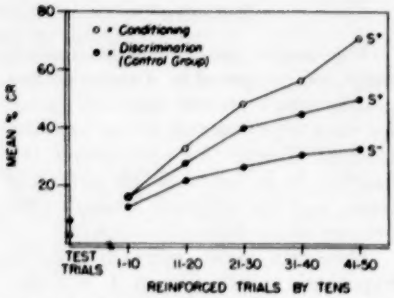

FIG. 1. Mean percentage of CR's in successive blocks of 10 reinforced trials for Group 1 and the Control Group.

sponses to CS + and CS − is only 17%. The explanation for the relatively poor discrimination under these conditions follows from the same major assumption as led to the prediction of a lowered level of responding to CS +, namely, that there was considerable generalization between CS + and CS − because of their similarity. This assumption is supported indirectly by a correlation of + .88 between total responses to CS + and CS − for each *S*. Evidently, when stimuli are similar, as in this experiment, high responsivity to CS + is likely to be accompanied by high responsivity to CS − (and vice versa).

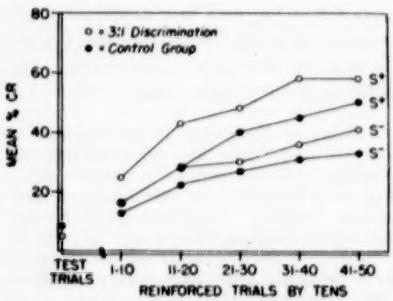

FIG. 2. Mean percentage of CR's in successive blocks of 10 reinforced trials for Group 2 (3:1 discrimination) and the Control Group.

Learning curves for Group 2 and the Control Group appear in Fig. 2. Inspection of the curves indicates that the predictions for this group—higher responding to both stimuli than the Control Group, yet no better discrimination—are supported. That discrimination is equivalent to that accomplished by the Control Group is readily shown by comparing the differences between mean percentage CR's to CS + and CS − on the last block of trials: 3:1 group, 17.5%, and Control Group, 17%. Analysis of variance supports the prediction of higher responding to both stimuli by the 3:1 group. The *F* for the effect of experimental condition was 3.34 which for 1 and 40 *df* is significant at the .08 level. Of interest is another finding: *F* for the effect of sex was 4.28 ($P < .05$, 1 and 40 *df*), indicating that women tend to respond at a higher level than men.

Figure 3 presents learning curves for Group 3 which received 20 preconditioning trials prior to discrimination training. Inspection of the curves shows that the predictions derived from Hull's theory are supported. First, it is clear that discrimination training produced an initial response decrement to CS +. Second, the level of responding to CS + and CS − is at a similar level

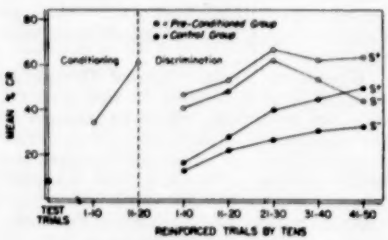

FIG. 3. Mean percentage of CR's in successive blocks of 10 reinforced trials for Group 3 (preconditioned discrimination) and the Control Group.

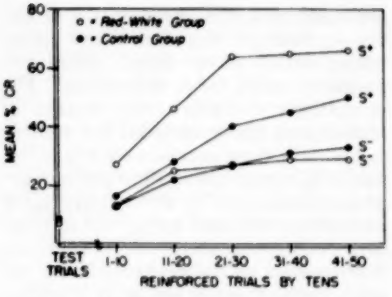

FIG. 4. Mean percentage of CR's in successive blocks of 10 reinforced trials for Group 4 (red-white discrimination) and the Control Group.

early in training, as would be expected on the basis of stimulus similarity and a consideration of the role of context stimuli. Finally, the two curves, as predicted, increase together before discrimination begins to develop. These interpretations are confirmed by statistical analyses. The difference between the mean number of responses to CS + on Conditioning Trials 11–20 and on the first block of discrimination trials is significant ($t = 2.67$, $P < .02$). A comparison of responses to CS + and CS − over the first two blocks of discrimination training produced a nonsignificant t of 1.09 indicating that the hypothesis of differences in response level early in discrimination training may be rejected. A t test for related measures comparing the number of responses to both stimuli on Discrimination Trials 1–10 with those given on Discrimination Trials 21–30 produced a t of 2.94 ($P < .02$). The difference in responding to CS + and CS − late in training (last two blocks of trials) was significant ($t = 3.40$, $P < .01$).

The data most pertinent to the hypotheses being investigated with Group 4, where there is an increased psychophysical difference between CS + and CS −, are the learning curves presented in Fig. 4. The curves suggest that under these conditions there occurred a higher level of response to CS + and better discrimination between CS + and CS − than in the Control Group. For the last block of trials, the difference between the difference scores (CS +–CS −) for these two groups is significant ($t = 2.58$, $P < .02$), thus supporting the interpretation that Group 4 discriminates better than the Control Group. The responses to CS + given by Group 4 and the Control Group over all trials were significantly different ($t = 2.20$, $P < .05$), lending support to the hypothesis that increasing stimulus dissimilarity produces a higher level of responding to CS +.

Learning curves for Group 5 are presented in Fig. 5. Inspection of these curves suggests that one may accept the predictions of lowered responding to CS + and no discrimination for the Partial Reinforcement Group. On the final block of trials, the mean percentage of CR's to CS + is identical with that to CS −. Response to CS + is lower for the Partial Reinforcement Group than for the Control Group on the last three

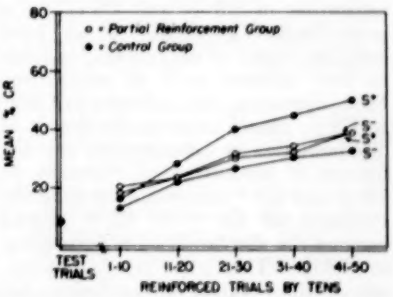

FIG. 5. Mean percentage of CR's in successive blocks of 10 reinforced trials for Group 5 (partial reinforcement) and the Control Group.

blocks of trials, but the difference is not reliable ($F = 1.70$, $P = .20$). It is possible that the reason for this is that CS + and CS − in the Control Group are already so similar that a further lowering of performance would be difficult to produce.

Discussion

The results of this experiment offer confirmation, in considerable detail, of the predictions derived from a classical S-R theory of discrimination learning. According to this theory, strength of conditioning and physical similarity between stimuli are two of the most important variables controlling differential conditioning. The relational point of view would make similar predictions concerning the effect of stimulus similarity on ease or difficulty of discrimination. However, such theorists would have nothing to say as to the relation of strengths of excitatory and inhibitory tendencies to differentiation of stimuli. Thus, the learning-theory framework permits more predictive precision for determining the influence on discrimination of the variables actually studied in this experiment.

Of the two sets of variables, the effects produced by manipulating the similarity of CS + to CS − are more straightforward and easier to handle. In theoretical terms, the effect of manipulating similarity is to control the degree of separation of CS + and CS − on a generalization gradient. As we have seen, the degree of such separation leads to two different sorts of predictions. First, decreasing the similarity of CS + to CS − should increase the degree of discrimination as measured by the difference in percentage of responses to CS + and CS −; second, decreasing the similarity of CS + to CS − should increase the absolute level of responding to CS +. This experiment has employed three values of similarity: (*a*) partial reinforcement where there is no difference between CS + and CS −, (*b*) a control group where there is one difference (position) between CS + and CS −, and (*c*) the Red-White group where there were three differences, position, color, and brightness. The percentages by which the number of responses to CS + exceeded the number to CS − in these groups were 0%, 17%, and 37%, respectively. The percentages of responses to CS + at the same point were 39%, 50%, and 66%. All of these data thus support theoretical predictions. With respect to the performance of the partial-reinforcement group, it is probably worth while pointing out that, contrary to what seems to be a general opinion (**6**), its performance level is considerably lower ($P < .01$) than that of a straight conditioning group (Group 1).

It is probably a matter of some importance that, even under the most effective conditions for the development of discrimination, a considerable percentage of responding to CS − persists. It may be that this is peculiar to the short-trace classical conditioning procedure, for by other methods (**5**) lower animals eventually develop perfect discriminations. If this result is peculiar to the experimental situation employed, it seems probable that the short interstimulus interval is the most important determiner. In this situation, there is probably not time for differentiating verbal reactions to occur. This would imply that the use of the short interval, in effect, increases similarity.

The influence of manipulating the strength of conditioning to CS + depends to a very important degree upon the method by which differences in this variable are achieved. If the strength of conditioning to CS + is increased by using a larger fraction of CS + trials than CS − trials, and if this procedure is employed from the beginning of training, the effect should be to increase the level of responding to both CS + and CS −, without influencing the amount of differential responding. The two different conditions employed in this experiment to manipulate this variable (Control Group and Group 2) produced results in line with theoretical pre-

dictions. An important question concerns the generality of this effect. A fairly obvious next step is to manipulate this variable through a larger range of values, including ratios where CS − trials are more numerous than CS + trials.

The preconditioning procedure for increasing the strength of conditioning to CS + leads to a different prediction. If such conditioning is carried to the limit of learning, the most likely effect of the introduction of CS − trials (i.e., when discrimination begins) is to reduce the level of responding to CS +, as a result of generalized inhibition. Beyond the initial trials, discrimination should begin to develop. The level of performance ultimately attained by the curves for CS + and CS − should be the same as those for another group, such as the control group in this experiment, where the practice conditions were otherwise identical. Inspection of Fig. 3 suggests that this outcome is possible; but unfortunately 100 trials were not enough for the curves, with the possible exception of the CS + curve, for the preconditioned group to reach final levels. This again suggests a logical next study in which the course of discrimination learning, following preconditioning, would be followed to the limit of the discrimination function.

Summary

This experiment studied the relation between increasing the difference in strength of conditioning to the stimuli and manipulating the physical differences between stimuli in the differential eyelid conditioning situation. There were 100 *Ss*, 14 in each of five different experimental conditions and 30 in a control condition. Group 1 consisted of *Ss* conditioned by the classical conditioning method. Comparison of this group with the Control Group, trained by the standard discrimination-learning procedure, produced as predicted a higher conditioning level of performance by Group 1 than by the Control Group ($P = .01$), although discrimination was evidenced by the Control Group ($P < .001$).

Experimental Group 2 received discrimination training in which the ratio of CS + to CS − was 3:1. Compared with the Control Group, Group 2 (*a*) responded at a higher level to both stimuli ($P = .08$), and (*b*) showed no improvement in discrimination.

Experimental Group 3 was conditioned to a criterion prior to receiving discrimination training. Results showed: (*a*) a decrement in performance immediately after introduction of discrimination training ($P < .02$), (*b*) no difference between level of response to CS + and CS − early in training, (*c*) an increase in responding to both stimuli prior to the development of discrimination ($P < .02$), and (*d*) better performance to CS + than to CS − late in training ($P < .01$).

Experimental Group 4 received discrimination training with increased stimulus differences. Compared to the Control Group, the *Ss* in this condition (*a*) discriminated better ($P < .02$), and (*b*) responded at a higher level to CS + ($P < .05$).

Experimental Group 5 received training under a partial-reinforcement schedule. Results showed that *Ss* in this group (*a*) were unable to discriminate, and (*b*) showed an unreliable trend toward responding at a lower level to CS + than the Control Group ($P = .20$).

These results were interpreted within the framework of Spence's theory of discrimination learning.

References

1. Antoinetti, J. A. The effect of discrimination training upon generalization. Unpublished Ph.D. dissertation, Yale Univer., 1950.
2. Ehrenfreund, D. An experimental test of the continuity theory of discrimination with pattern vision. *J. comp. Psychol.*, 1948, **41,** 408–422.
3. Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43,** 313–320.
4. Grice, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38,** 633–642.
5. Harlow, H. F. Studies in discrimination learning by monkeys: I. The learning of discrimination series and the reversal of discrimination series. *J. gen. Psychol.*, 1944, **30,** 3–12.
6. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

7. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
8. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25,** 141–158.
9. Kendler, T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40,** 552–562.
10. Kimble, G. A., & Dufort, R. H. The associative factor in eyelid conditioning. *J. exp. Psychol.*, 1956, **52,** 387–392.
11. Kimble, G. A., Mann, L. I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49,** 407–417.
12. Margolius, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49,** 105–111.
13. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.
14. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43,** 427–449.
15. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44,** 430–444.
16. Spence, K. W. Failure of transposition in size discrimination of chimpanzees. *Amer. J. Psychol.*, 1941, **54,** 223–229.
17. Spence, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31,** 257–271.
18. Spence, K. W., & Beecroft, R. G. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48,** 399–403.
19. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47,** 127–134.
20. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41,** 81–92.

(Received June 14, 1956)

Journal of Experimental Psychology
Vol. 53, No. 6, 1957

CONTINGENT PARTIAL REINFORCEMENT AND THE ANTICIPATION OF CORRECT ALTERNATIVES[1]

HOWARD BRAND, JAMES M. SAKODA, AND PAUL J. WOODS[2]

The University of Connecticut

Estes and Straughan (**6**) developed an equation for the two-choice noncontingent reinforcement situation. Here S has two alternative responses either one of which may be followed by one of two environmental events. The occurrence of the events is controlled entirely by E and is independent of S's choices. A typical situation is one where there are two levers (alternative responses) with a light over each one (environmental events); S is to predict which light is going to flash next by pressing one of the levers. After S has pressed one of the levers E may then flash either the left or the right light according to a predetermined probability schedule.

Contingent reinforcement, however, differs from noncontingent in that the events following the responses are, in part, S-controlled. Typically S has two alternative responses, and one of two environmental events may subsequently occur, but they are not independent of S's responding. In such a situation, there may be two levers and a ground-glass screen behind which are red and white bulbs. The levers are controlled by E so that either one can switch on either light and are manipulated according to a predetermined probability schedule. The S is to predict which lever will give, say, a red light. If E has set the switches so that the left will give white and the right will give red, and S presses the right lever, he will "get" a red light. Here, then, the environmental event depends upon two probabilities: the probability of the red light being assigned to the right lever (controlled by E) and the probability of the right lever being pressed (controlled by S and the previous sequence of trials).

In the present study, an equation is used to predict response probability $\tilde{p}(n)$, in a two-choice contingent reinforcement situation, and it is tested in several experiments. The equation was derived by following analogously Estes and Straughan's (**6**) derivation for noncontingent reinforcement.[3]

Let A_1 and A_2 be two disjunctive response classes and let E_1 and E_2 represent reinforcing stimuli (environmental events). The S responds with either A_1 or A_2 and either E_1 or E_2 follows. The probability of an E_1 following an A_1 is π_1 (i.e. experimenter-controlled probability of red light on right lever), of an E_2 following an A_1 is $1 - \pi_1$ (probability of white light on right lever), of an E_1 following an A_2 is π_2 (probability of red light on left lever), and of an E_2 following an A_2 is $1 - \pi_2$ (probability of white light on left lever). Further, let $p(n)$ denote the probability of A_1 after Trial n (i.e., the probability of S pressing the right lever and $1 - p(n)$ the probability of A_2 after Trial n (i.e., the probability of S pressing the left lever). Finally, let θ equal the proportion of stimulus elements sampled on each trial.

To utilize Estes' (**4**) model for contingent reinforcement it is necessary to assume that accompanying an overt $A_2 - E_2$ sequence there is a covert $A_1 - E_1$ sequence and the sampled stimulus elements present are conditioned to A_1. (This is saying that when S gets a white light

[1] The present research was conducted under grant M-814C from the National Institute of Mental Health.

[2] The collection and analysis of data was mainly done by Paul J. Woods. Howard Brand and James M. Sakoda are senior staff on the project.

[3] This equation was derived by Paul J. Woods with the aid of David Zeaman.

he assumes that he could have gotton a red light had he pressed the other lever.)

The change in response probability for the contingent situation is obtained from the rules given by Estes and Straughan (**6**) in the following way. On an $A_1 - E_1$ sequence and on an $A_2 - E_2$ sequence

$$p(n+1) = (1-\theta)p(n) + \theta \qquad (1)$$

while on an $A_1 - E_2$ sequence and on an $A_2 - E_1$ sequence

$$p(n+1) = (1-\theta)p(n). \qquad (2)$$

In randomized contingent reinforcement Equation 1 will be applicable in the proportion π_1 times the probability $p(n)$ and in the proportion $(1-\pi_2)$ times the probability $1-p(n)$. Equation 2 will be applicable in the proportion $1-\pi_1$ times the probability $p(n)$ and in the proportion π_2 times the probability $1-p(n)$.

The average probability of A_1 after Trial $n+1$ may now be written:

$$\begin{aligned} \bar{p}(n+1) = {} & \bar{p}(n)\pi_1[(1-\theta)\bar{p}(n)+\theta] \\ & + \bar{p}(n)(1-\pi_1)[(1-\theta)\bar{p}(n)] \\ & + [1-\bar{p}(n)]\pi_2[(1-\theta)\bar{p}(n)] \\ & + [1-\bar{p}(n)](1-\pi_2) \\ & \times [(1-\theta)\bar{p}(n)+\theta]. \end{aligned} \qquad (3)$$

By rearrangement and induction Equation 3 is put into the general equation which predicts the probability of A_1 after Trial n:

$$\bar{p}(n) = \frac{1-\pi_2}{2-\pi_1-\pi_2} - \left[\frac{1-\pi_2}{2-\pi_1-\pi_2} - \bar{p}(O)\right] \times [1-\theta(2-\pi_1-\pi_2)]^n. \qquad (4)$$

This equation, derived from basic assumptions in Estes' theory, is identical with one derived by Bush and Mosteller from different assumptions (**2,** p. 288). It is also comparable to an equation derived by Detambel (**3,** p. 99). His situation is more general in that he deals with any number of response alternatives, but the equation he derives is for the special case where π_1 equals a finite probability and the reward probability on all other alternatives equals zero.

Equation 4 obviously differs from the one for the noncontingent reinforcement situation. Estes and Straughan (**6**) give the following equation:

$$\bar{p}(n) = \pi - [\pi - \bar{p}(O)](1-\theta)^n.$$

The first term on the right in each equation is the predicted final response level of the asymptote. (That is, as n becomes large in each equation the last term on the right approaches zero and $p(n)$ approaches the first term on the right as a limit.) The asymptote in the noncontingent situation depends upon the probability of being rewarded for an A_1 response, π. (The probability of reward for the alternative response must equal $1-\pi$ since with noncontingent reinforcement the probabilities must sum to unity.) The predicted asymptote in the contingent situation depends upon the probabilities of reward on *both* alternatives. (It is not necessary that they sum to unity in the contingent situation. In fact, if they do sum to unity the contingent prediction equation degenerates to the noncontingent form.)

The rate constant, the term which determines the rate of rise of the predicted curve, in the noncontingent situation is $(1-\theta)$; it is independent of any reward probabilities. In the contingent situation, however, the rate constant, $[1-\theta(2-\pi_1-\pi_2)]$, is dependent upon reward probabilities as well as θ.

The main purpose of the following experiments was to test Equation 4.

Method

Design.—The 22 reinforcement schedules used are shown in Table 1. In selecting these schedules, two variables were used: the difference and the ratio between the probabilities of the correct outcome. For instance, a 60:30 schedule means that for A_1 responses, E_1's will occur 60% of the time while for A_2 responses, E_1's will occur 30% of the time. This schedule has a probability difference of 30 and a probability ratio of 2:1. The schedules varied from a probability difference of 20 to 50 and from a probability ratio of 2:1 to 11:1. Each condition had six *S*s assigned to it, and *S*s were given 120 acquisition trials (**1**).

Subjects.—The *S*s were 132 undergraduate volunteers from courses in psychology at the University of Connecticut.

Apparatus.—The apparatus consisted of two levers (microswitches) at table height in front of *S* and a vertical panel containing a ground-glass screen at eye level and a green signal light. Red and white bulbs were enclosed in a box behind the vertical panel and flashed on the ground-glass screen when *S* pressed the levers. The levers were controlled independently by *E* through a switching set-up so that each lever could give either color. By instruction, the red light was selected as the reward. The reinforcement was contingent upon *S*'s choice of a lever following a given probability schedule.

Switches on *E*'s control panel enabled him to set up one of four possible situations before each response of *S*. That is, the apparatus could be set to give red on the left lever and white on the right lever, or the reverse of this, or red on both levers, or white on both levers. When *E* had

TABLE 1

VALUES OF THEORETICAL AND EMPIRICAL ASYMPTOTES AND $\theta(2 - \pi_1 - \pi_2)$ FOR RATIOS AND DIFFERENCES FOR EACH SCHEDULE

Condition		Asymptotes		$\theta(2 - \pi_1 - \pi_2)$		
				Theoretical**		Empirical
Difference	Ratio	Theoretical	Empirical	Block*	Value	Value
50	100:50 (2:1)	1.00	.917	12	.0275	.0608
47	94:41 (2:1)	.898	.812	12	.0193	.0405
45	90:15 (2:1)	.846	.821	12	.0251	.0307
43	86:43 (2:1)	.803	.796	12	.0501	.0548
40	80:40 (2:1)	.750	.767	12	.0440	.0221
30	60:30 (2:1)	.636	.562	12	.0090	.0245
20	40:20 (2:1)	.571	.631	12	.0000	.0026
50	95:45 (2.1:1)	.917	.896	12	.0227	.0253
50	90:40 (2.25:1)	.857	.816	12	.0282	.0505
50	83:33 (2.5:1)	.798	.871	12	.2489	.0271
50	75:25 (3:1)	.750	.833	6	.0792	.0308
40	60:20 (3:1)	.666	.750		1.000	.0329
30	45:15 (3:1)	.607	.546	12	.0022	.0046
20	30:10 (3:1)	.562	.667	5	.0150	.0166
50	60:10 (6:1)	.692	.750		1.00	.0735
40	48:8 (6:1)	.638	.738	3	.1682	.0594
30	36:6 (6:1)	.594	.650	9	.0596	.0172
20	24:4 (6:1)	.558	.575	12	.0362	.0239
50	55:5 (11:1)	.678	.763	4	.1092	.0344
40	44:4 (11:1)	.631	.733	12	.2799	.0294
30	33:3 (11:1)	.591	.604	6	.0941	.1171
20	22:2 (11:1)	.556	.630	6	.1295	.0212

* The number of empirical trial blocks used in estimates of $\theta(2 - \pi_1 - \pi_2)$.
** $\theta(2 - \pi_1 - \pi_2)$ calculated on basis of theoretical asymptote.

the switches set he flashed the green signal light on *S*'s panel, and *S* pressed a lever. A light on *E*'s control panel told him which lever *S* had pressed. The known factors were, therefore, the probability of a red light on each side according to the predetermined probability schedule, the lever that *S* pressed on each trial, and whether or not reward was obtained on each trial.

Instructions.—In the instructions an attempt was made to set up a situation for which the model would apply. That is, it was necessary to lead *S*s to believe that whenever they were incorrect they could have been correct had they chosen the other alternative. Obviously, this was not really the case.

After *S* was seated comfortably in front of the apparatus the following instructions were given:

"The two levers in front of you are merely switches which flash either red or white lights on the ground-glass screen. Your task is to try to get as many red lights as you can. But this lever (pressing the one set to give red) will not always be red; sometimes the other lever will be red. The red light will be manipulated back and forth from trial to trial according to a scheme or plan. On a particular trial, then, one lever will be correct and one incorrect. If you press the correct lever a red light will flash and if you press the incorrect lever a white light will flash. Your task then is to try to figure out the scheme or plan so that on each trial you can press the one lever that will give you a red light.

"When the green signal light flashes you are to press one lever and one lever only. Then wait; the green signal light will flash again; this will be the next trial; you then press the lever that you think will be red this time. Remember, each trial you press only one lever, and you press the lever that you think will give you a red light.

"We can't stop once we have started, so do you have any questions now?"

The *S*s were led to believe that there was a scheme or plan to the randomized schedule of

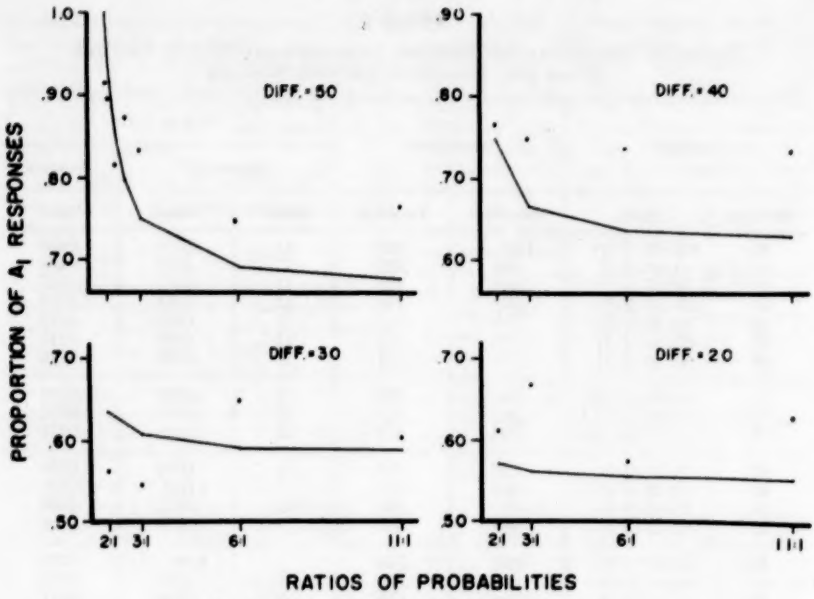

FIG. 1. The predicted and empirical asymptotes.

rewards, following the convention established by Estes and Straughan (6).

RESULTS

Asymptotic levels.—The formula for the predicted asymptote in a contingent situation given by Estes (5) is

$$\frac{1 - \pi_2}{2 - \pi_1 - \pi_2},$$

where $\pi_1 > \pi_2$. In Equation 4 as n becomes large,

$$\bar{p}(n) \xrightarrow{\lim} \frac{1 - \pi_2}{2 - \pi_1 - \pi_2}. \qquad (5)$$

The proportions of responses to the most profitable side on the final third of the acquisition trials were used for empirical asymptotes. Predicted and empirical asymptotes are shown in Table 1. In the main, the empirical asymptotes are larger than the theoretical estimates. Some of the error is undoubtedly due to the fact that the empirical asymptotes are based upon the mean curves of only 6 Ss. Nevertheless, errors should be expected in both directions and not predominantly in just one.

In Fig. 1 the predicted and empirical asymptotes are plotted for each of the probability differences separately with ratios of probabilities on the abscissa and the proportion of A_1 responses on the ordinate.[4] The predicted values are connected by the solid line and the empirical points are placed singly. The figure clearly shows that the theoretical values are predominantly underestimates of the empirical values.

The results of an analysis of variance to test the significance of

[4] The asymptotic results for three of the schedules (94:47, 90:45, and 86:43) are not included in Fig. 1 since they have unique differences.

TABLE 2

ANALYSIS OF VARIANCE OF THE DEPARTURE OF EMPIRICAL ASYMPTOTES FROM THEORETICAL VALUES

Source	*df*	*MS*	*F*
Total deviations from expected	22	268.94	2.04**
Mean level of deviations	1	714.94	5.42*
Variability in deviations	21	247.70	1.88*
Within (error)	110	132.01	
Total	132		

*$P = .05$.
**$P = .01$.

departure from theoretical values are shown in Table 2. The total deviation from expected values was significant ($P < .01$). Partition of the sum of squares into mean level of deviations and variability in deviations obtains mean squares significant at the .05 level. The significance of the mean level of deviations indicates that the discrepancy between the predicted and empirical asymptotes tended in one direction and did not fluctuate equally in both directions. And, since the variability in deviations was significant, it can also be concluded that the size of the departure from the theoretical values was not the same for the different schedules.

For the 16 schedules indicated in Table 3 another analysis of variance was carried out to assess the influence of difference of probability of reward and ratio of probability of reward on these departures from theoretical values. For this purpose, the sums of squares for variability in deviations contributed by these 16 schedules were calculated. These results are shown in Table 3. Using the within mean square obtained from the analysis of all 22 schedules as the best estimate of error it is found that both differences and ratios contribute significantly to the departure from theoretical values. When the effect of each difference and each ratio is analyzed separately there is a trend for the mean squares of the ratios to increase as the ratios increase in size. No such consistent trend is found for the differences.

Rate constants.—Equation 4 may be transformed to describe the obtained data. Let m designate the ordinal number of a block of 10 trials which runs from Trial $n + 1$ to Trial $n + 10$, and let $\bar{P}(m)$ represent the expected proportion of A_1 responses in the block. Also, to make the expression easier to handle let

$$\frac{1 - \pi_2}{2 - \pi_1 - \pi_2} = \alpha$$

and

$$2 - \pi_1 - \pi_2 = a.$$

It can then be shown that

$$\bar{P}(m) = \alpha - \frac{[\alpha - \bar{p}(O)](1 - \theta a)^{10(m-1)}}{10\theta a} \times [1 - (1 - \theta a)^{10}]. \quad (6)$$

TABLE 3

ANALYSIS OF VARIANCE SHOWING THE INFLUENCE OF DIFFERENCES AND RATIOS ON THE DEPARTURE OF EMPIRICAL ASYMPTOTES FROM THEORETICAL VALUES

Source	*df*	*MS*	*F*
Ratios (R)	3	435.40	3.30*
(2:1)	1	399.91	3.03
(3:1)	1	667.82	5.06*
(6:1)	1	800.42	6.06*
(11:1)	1	1126.14	8.53**
Differences (D)	3	388.88	2.94*
(50)	1	205.93	1.56
(40)	1	1377.14	10.43**
(30)	1	65.34	.49
(20)	1	849.66	6.44*
R × D	9	120.59	.91
Within	110	132.01	

*$P = .05$.
**$P = .01$.

This expression is the mean value of $\bar{p}(n)$ over the mth block of 10 trials. $\bar{P}(1)$ can be measured more accurately than $\bar{p}(O)$ and hence it would be desirable to transform our equation so that $\bar{P}(1)$ can be used instead of $\bar{p}(O)$. Following Estes and Straughan (6), Equation 6 for $m = 1$ may be written

$$\bar{P}(1) = \alpha - \frac{[\alpha - \bar{p}(O)][1 - (1 - \theta a)^{10}]}{10\theta a};$$

then solving for $[\alpha - \bar{p}(O)]$,

$$[\alpha - \bar{p}(O)] = \frac{10\theta a[\alpha - \bar{P}(1)]}{1 - (1 - \theta a)^{10}};$$

then substituting this into Equation 6 one obtains

$$\bar{P}(m) = \alpha - [\alpha - \bar{P}(1)] \times (1 - \theta a)^{10(m-1)}. \quad (7)$$

The empirical estimate of θa, still following Estes and Straughan (6), is obtained by summing Equation 7 over

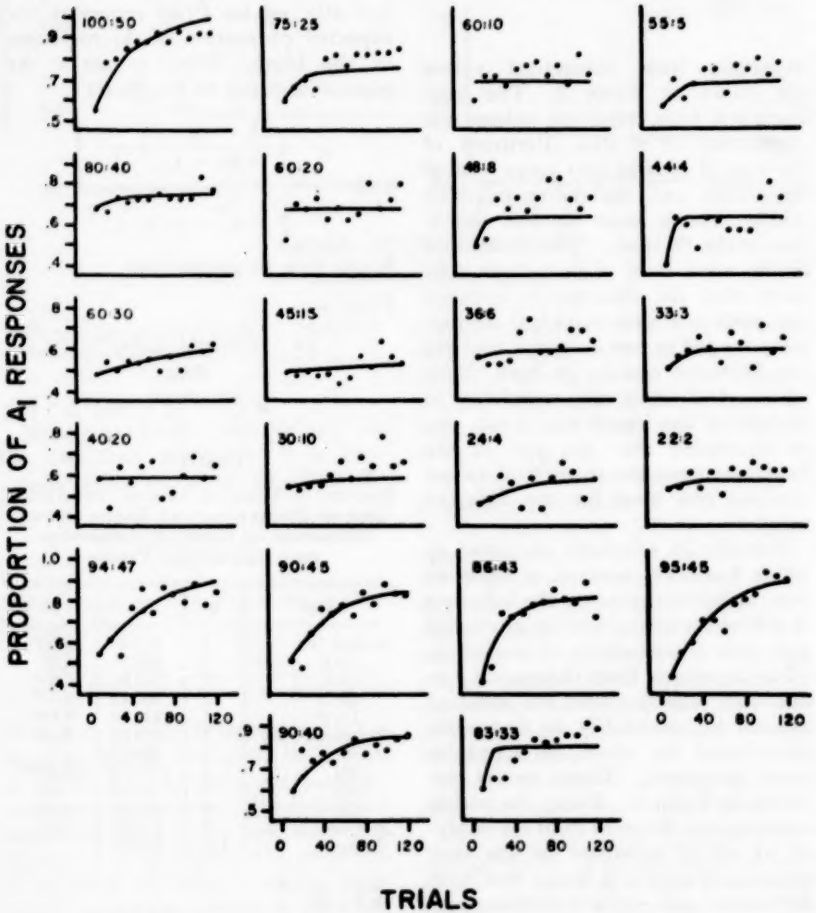

FIG. 2. The empirical data and the predicted curves for each schedule.

all values of m, so that for K blocks of trials one obtains

$$\sum_{m=1}^{K}\{\alpha - [\alpha - \bar{P}(1)](1 - \theta a)^{10(m-1)}\} = K\alpha - [\alpha - \bar{P}(1)] \times \frac{1 - (1 - \theta a)^{10K}}{1 - (1 - \theta a)^{10}}. \quad (8)$$

Equation 8 is then equated to the sum of the observed proportions, $\bar{P}$, for a given series, and solved for θa by using successive approximations.

The values obtained in this manner are shown in Table 1. The θ's for two of the schedules were assigned maximum values of 1.0 since Equation 8 could not be solved even when only the first two trial blocks were used. This value yields the best curve the theory can predict for the results obtained. The theoretical curves based on these values and calculated by means of Equation 7 are shown in Fig. 2. These findings clearly show that the predicted curves do not fit the data very well in many of the schedules.

Rate constants which better describe the empirical data can be obtained by substituting an empirical asymptote in Equation 8 rather than the theoretical asymptote and solving for θa on the basis of these values. The values of θa obtained in this manner are shown in Table 1. Curves could be drawn on the basis of these values that would fit the data quite well. These have not been included, however, since the purpose was to test a prediction equation derived from Estes' theory (**4**), and not to ascertain a curve of best fit for the present data.

Discussion

While some of the predicted curves are fairly good fits to the empirical points a large number of them are not. The discrepancy in some cases is obviously due to the fact that the assumption of a covert $A_1 - E_1$ sequence following an overt $A_2 - E_2$ sequence is not warranted. By the instructions an attempt was made to set up an experimental situation for which this condition would be satisfied. But for the low-rewarding schedules in particular, one can almost expect that it would be impossible to establish this condition by instructions. Take a 30:10 schedule, for example. Suppose S is responding 50% of the time to each alternative; 30% of his responses are rewarded at one alternative and 10% at the other which means that 20 times in 100 he is "correct" and 80 times in 100 he is "wrong." For the model to apply, S would have to assume that these 80 times in 100 he is "wrong" he could have been "correct" had he chosen the other alternative.[5]

The failure of the model to apply in some cases may be due to the fact then that the assumption of a covert correction response is not warranted. But there are quite probably other factors as well which contribute to the obtained discrepancies. Many of the empirical records do not reach an asymptote but still seem to be rising, which leads us to wonder if Ss might ultimately have gone to 100% responding on the most profitable side had they been run for enough trials. Flood (**7**) discusses this possibility in a different experimental context. He hypothesizes that the pure type of behavior (100% responses to the most profitable side) will be found in Ss who are convinced of randomness while the mixed type of behavior (trying to anticipate reward on the least profitable side as well) would be obtained when Ss believed there was some sort of pattern. In the present instructions to Ss they were told that the rewards were being manipulated according to a "scheme or plan." That is, a "pattern" set was induced. The rewards, however, were

[5] Bush and Mosteller (**2**) have suggested an alternative prediction equation in the context of their experimentation which does not make the equal alpha assumption; i.e., the reward parameter is not assumed to be equal to the nonreward parameter as it was in the derivation of Equation 4.

actually manipulated in a random fashion. It may turn out then that when the "pattern" set established by the instructions breaks down (as the present *S*s report) *S*s will gradually increase their proportion of responses to the most profitable side until they finally reach 100% responding to that side. It would be necessary to run more acquisition trials to test this possibility. It will be necessary to investigate the effects of instructional set in another experiment.

Summary

The purpose of the experiment was to test an application of Estes' statistical learning theory in a two-choice contingent partial-reinforcement situation. For the model to apply it is necessary for *S*s to assume that the alternative response would have been correct when they make an incorrect response. An attempt was made to establish this condition by means of instructions. Twenty-two different partial-reinforcement schedules were used, each one given to six *S*s. The schedules differed in terms of two variables: increasing difference and increasing ratio between the probability of the correct outcome. An analysis of variance shows that the asymptotes predicted by the theory on the basis of the experimental schedules significantly underestimate the empirical asymptotes.

Rate constants $(1 - \theta(2 - \pi_1 - \pi_2))$ were estimated from the empirical records with respect to the predicted asymptotes and with respect to the estimated empirical asymptotes. Theoretical curves were drawn to the predicted asymptotes. For some of the schedules the theoretical curves fit the data fairly well, but for others the fits are very poor. The contribution of several possible factors to these discrepancies were discussed.

References

1. Brand, H., Woods, P. J., & Sakoda, J. M. The anticipation of reward as a function of partial reinforcement. *J. exp. Psychol.*, 1956, **52**, 18–22.
2. Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.
3. Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–104.
4. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
5. Estes, W. K. Individual behavior in uncertain situations: an interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.) *Decision processes.* New York: Wiley, 1954.
6. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psyckol.*, 1954, **47**, 225–234.
7. Flood, M. M. Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.) *Decision processes.* New York: Wiley, 1954.

(Received June 8, 1956)